

Introduction to Medical Bioinformatics

Oct 2, 2018

Faculty of Medicine Siriraj Hospital

Mahidol University



SIRE 503: Next-generation sequencing technologies

Bhoom Suktitipat, MD, PhD
[bhoom.suk@mahidol.ac.th]

Department of Biochemistry
Faculty of Medicine Siriraj Hospital

&

Integrative Computational BioScience Center
Mahidol University

Outline

- Structure of the Nucleic Acid
- DNA Sequencing
- History of Next-generation Sequencing
- Template amplification strategy & Target enrichment
- Short-read vs Long-read sequencing methods
- Comparison
- Application
- Limitation
- Reference Genomes

Term Paper – Due Nov 15

Pick an original article from this list that is **related to bioinformatics data analysis**. Focus more on articles that use high-throughput data in their research, either genomics, transcriptomics, proteomics, metabolomics, or other big data related to biological questions and experiments. Stay away from papers that focus mainly on classical epidemiology, conventional risk factors, or biostatistics, that do not have much utilization of biological data.

- | | |
|--|---|
| 1. <u>Science</u> | 9. <u>Cell</u> |
| 2. <u>Nature</u> | 10. <u>PNAS</u> |
| 3. <u>Nature Genetics</u> | 11. <u>PLOS Genetics</u> |
| 4. <u>Nature Medicine</u> | 12. <u>PLOS Medicine</u> |
| 5. <u>Nature Cell Biology</u> | 13. <u>Genome Research</u> |
| 6. <u>New England Journal of Medicine</u> | 14. <u>Genome Biology</u> |
| 7. <u>British Medical Journal (BMJ)</u> | 15. <u>Human Mutation</u> |
| 8. <u>American Journal of Human Genetics</u> | 16. <u>Human Molecular Genetics</u> |

Have other ground breaking papers? Let's discuss!

Term Paper

Answer the following questions (3 sections)

Section 1: What is already known on this topic

In less than three single-sentence bullet points, please summarize the state of scientific knowledge on this topic. Emphasize on “**why**” this study needed to be done.

Section 2: What this study adds

In one or two single sentence bullet points, give a simple answer to the questions “**What do we now know as a result of this study that we did not know before?**” “**Is there any implications for practice, research, policy, or public health?**”

Be brief, succinct, specific, and accurate.

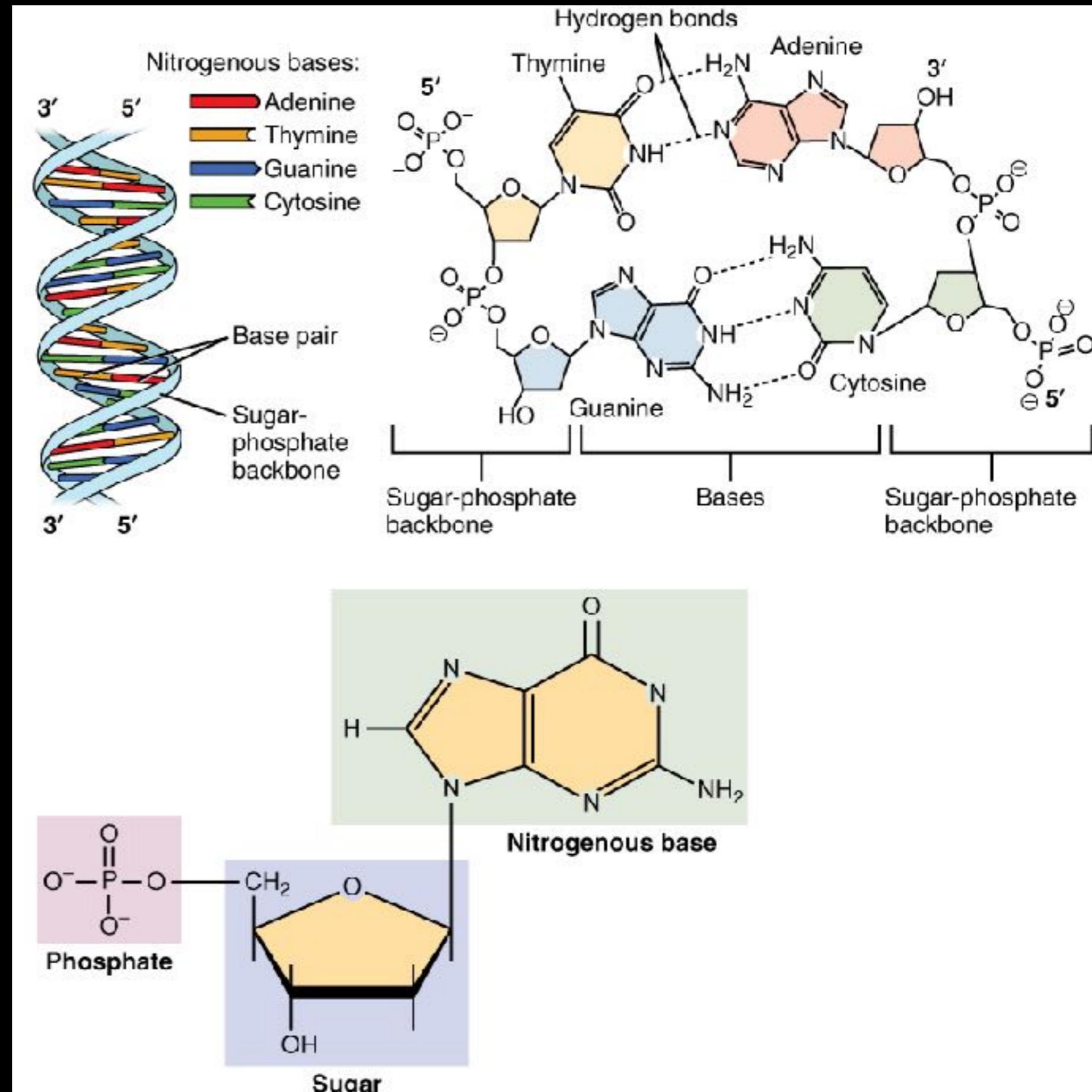
Term Paper

Section 3: How the data were analyzed

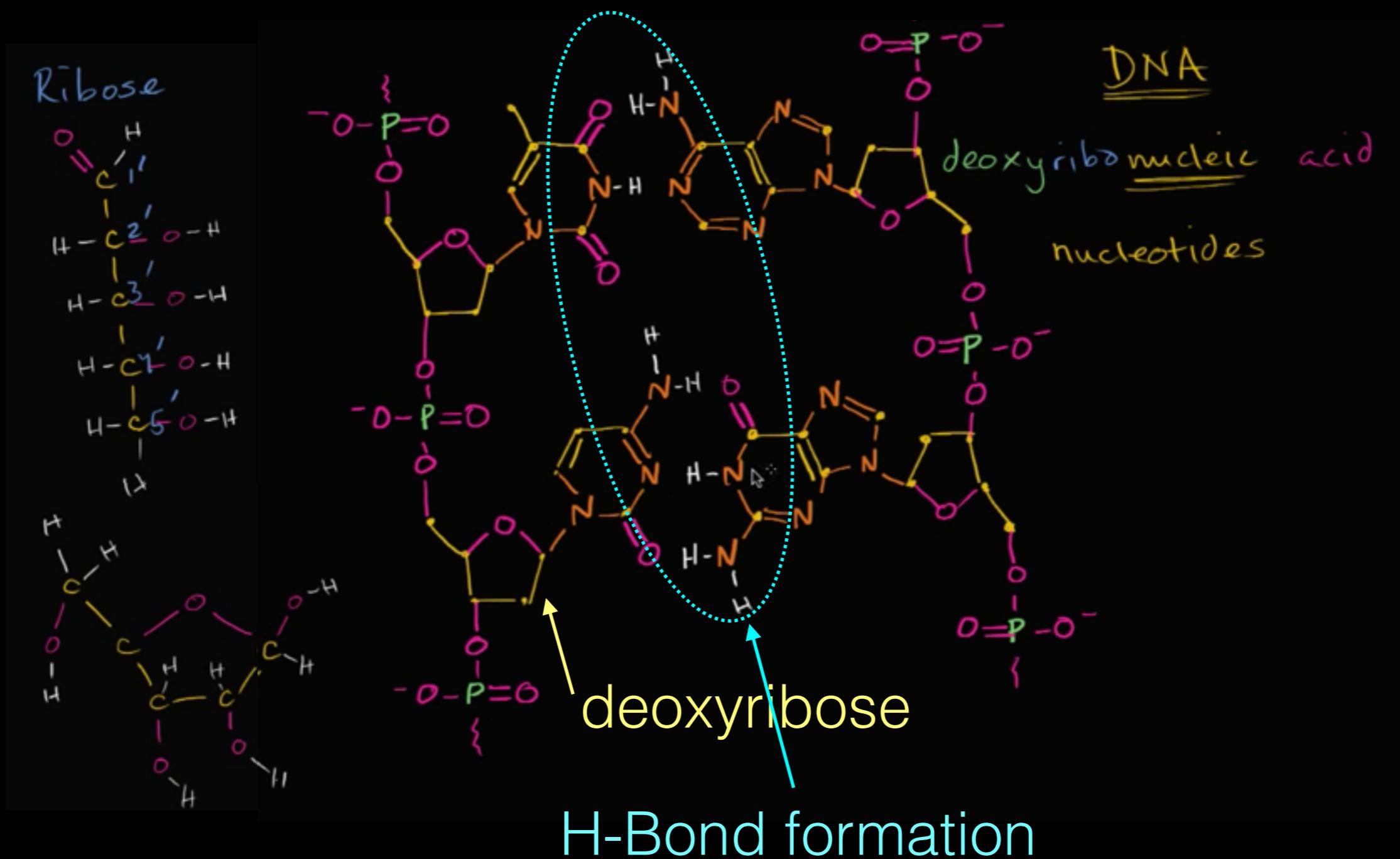
Pick 3 figures and describe the followings

1. What is question that this figure tried to answer?
2. What data have been generated to answer the question?
3. What analysis have been done to get to the conclusion?

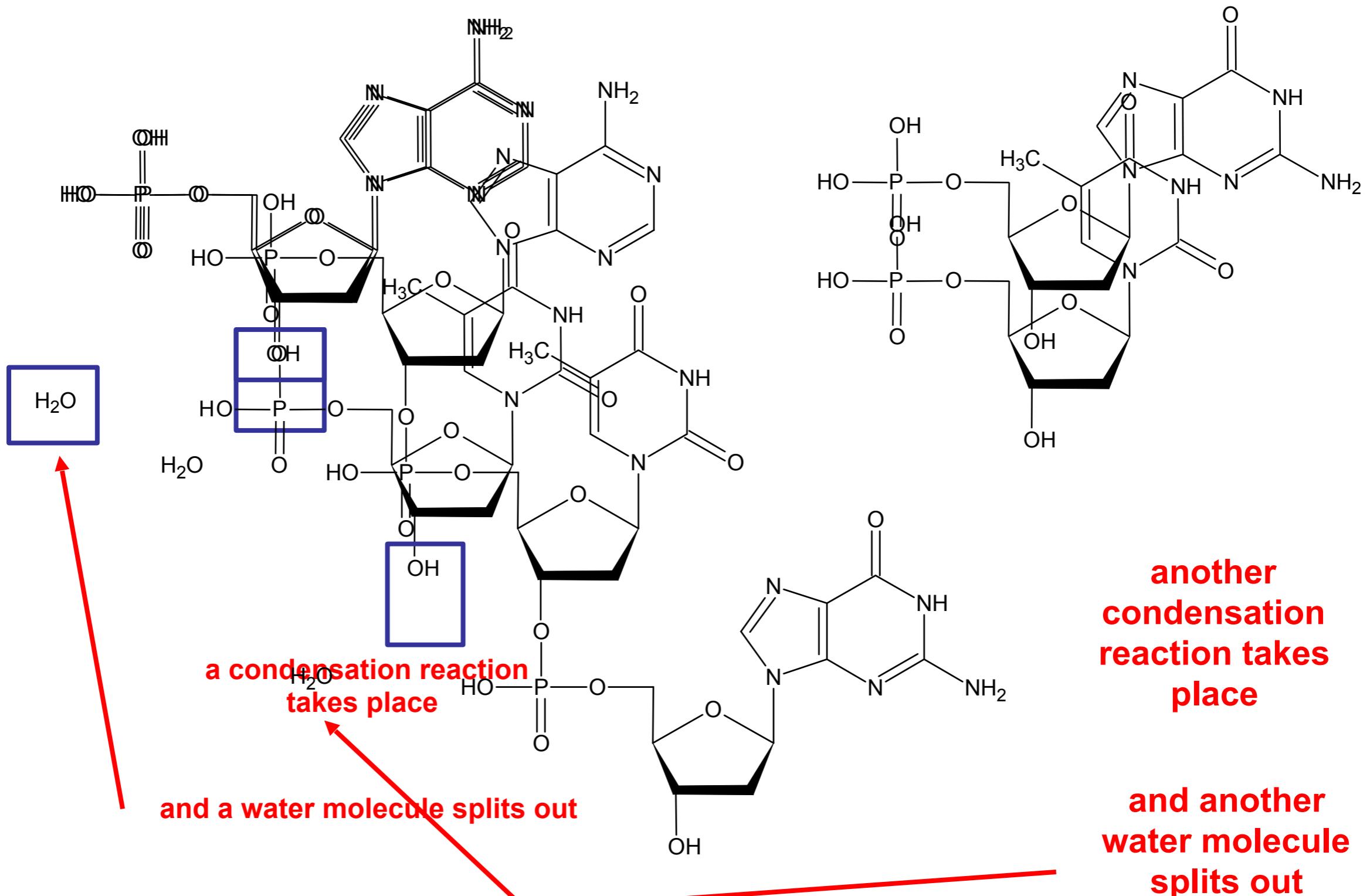
Structure of Nucleic Acid



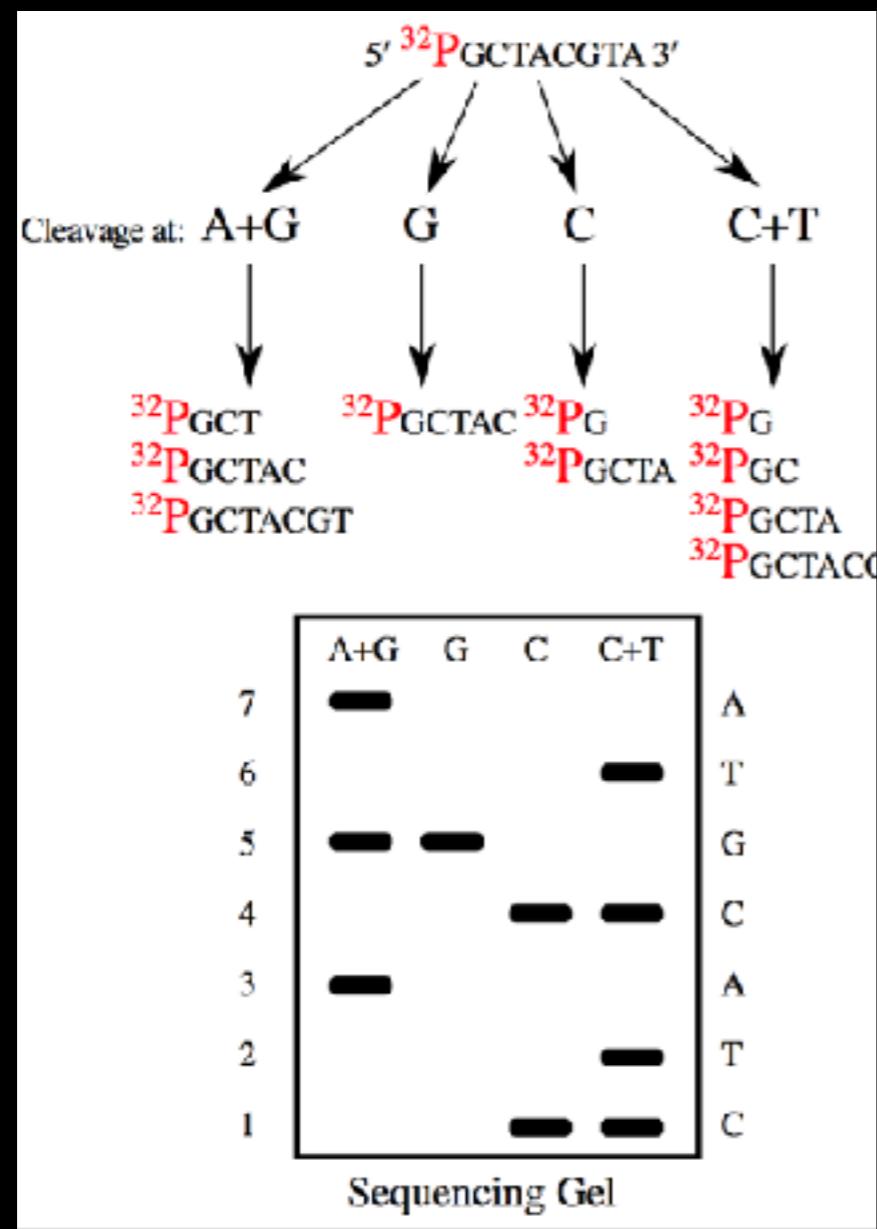
Chemical Structure of the DNA



Formation of a polynucleotide



DNA Sequencing

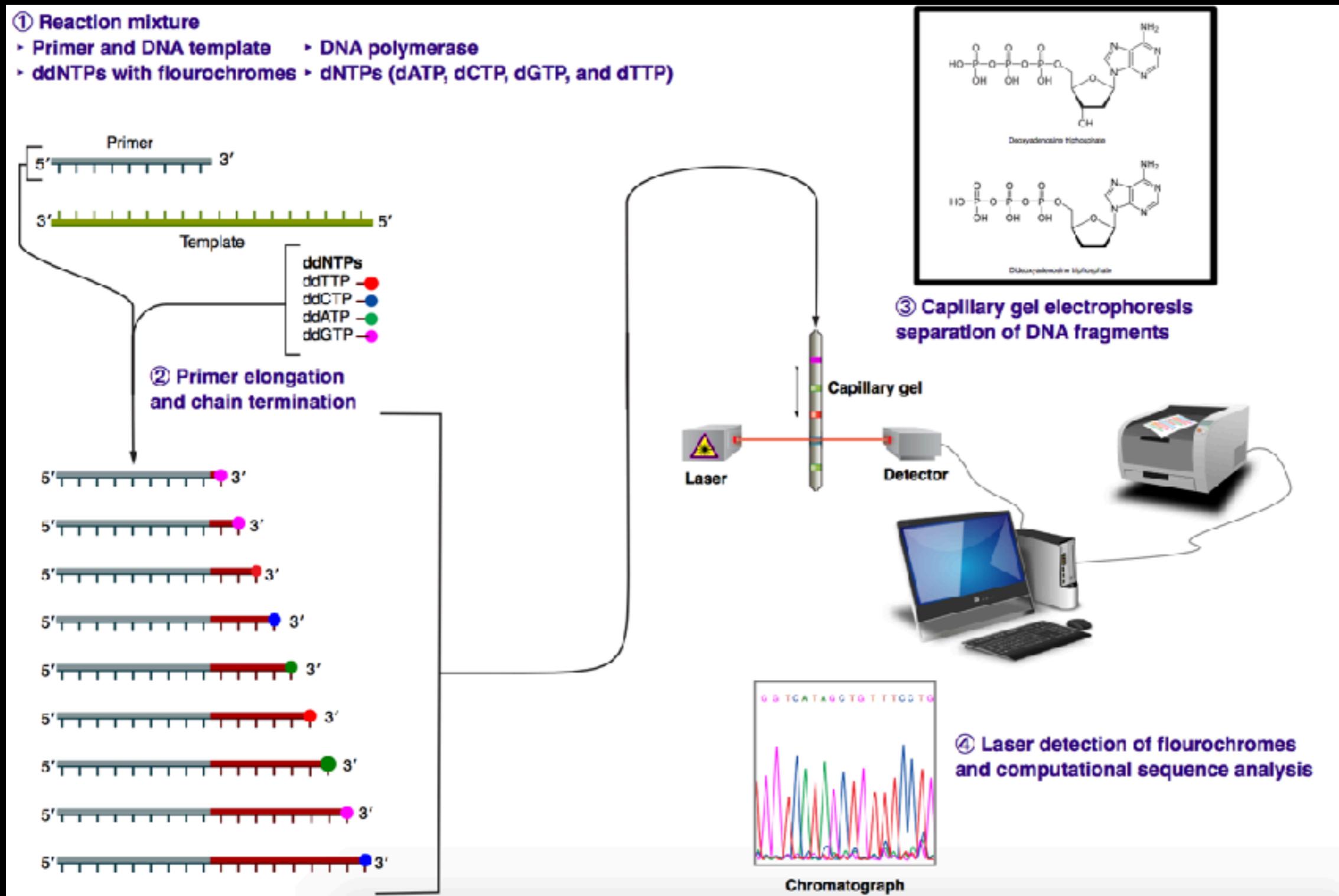


Initially popular, due to the ability to use the purified dsDNA directly.

- The 5' ends of the resulting fragments were then labeled with ^{32}P . These radio-labeled fragments were then aliquoted into four reactions: one fragmented by a chemical reaction that attacked only C nucleotides; another attacked only G nucleotides; a third cleaved A, with some cleavage of G nucleotides; and the fourth predominantly cleaved T nucleotides, with some cleavage at C sites.
- able to resolve homopolymer regions accurately with sequentially sized bands indicating all bases in the tract

**Nucleotide cleavage method
(AM Maxam and W Gilbert)**

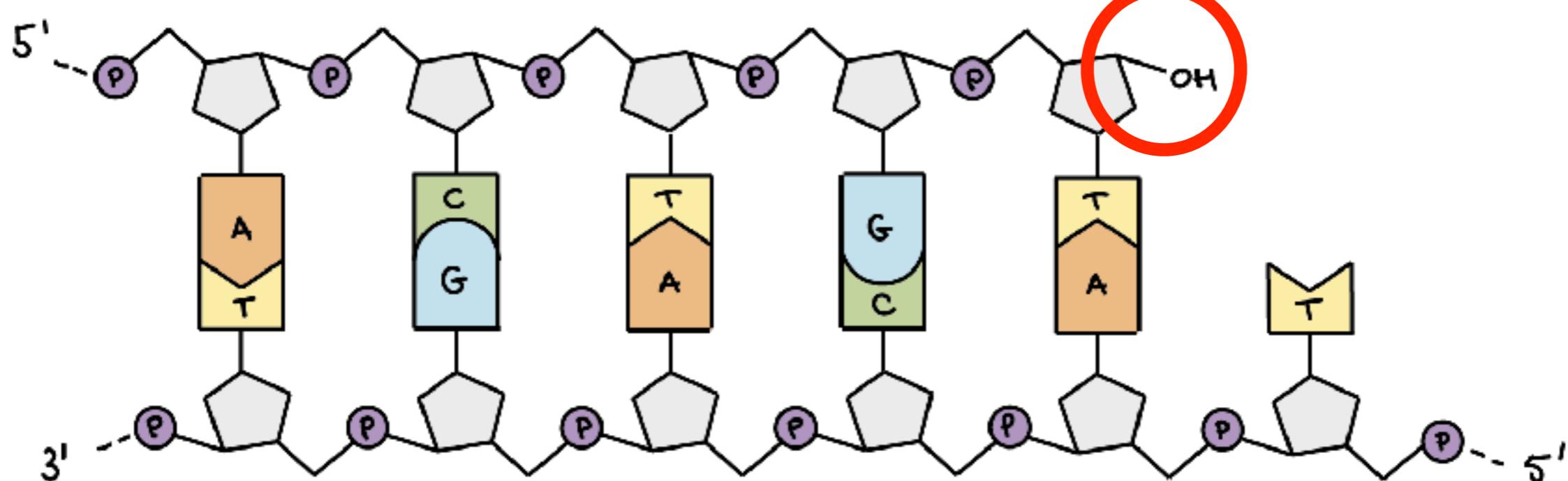
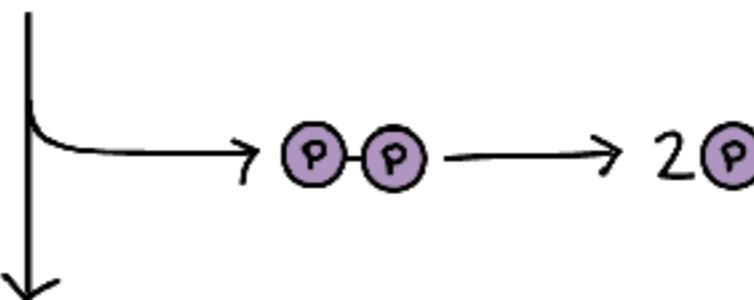
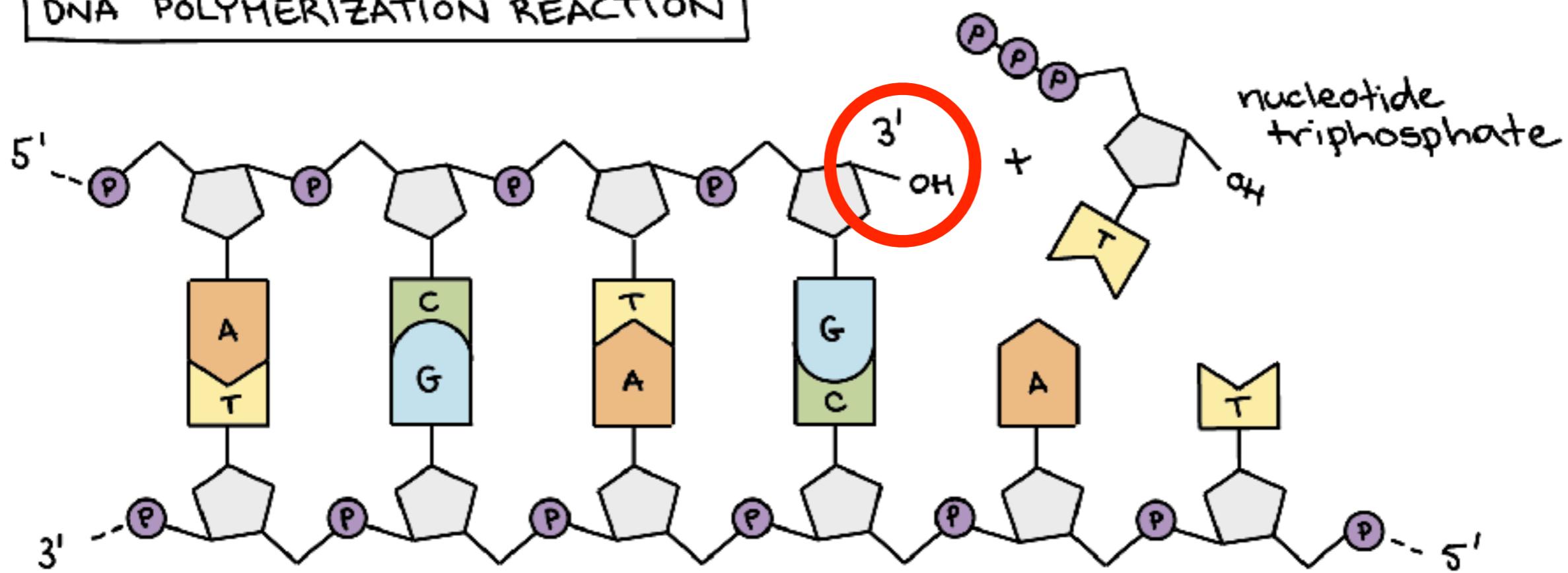
Sanger Sequencing



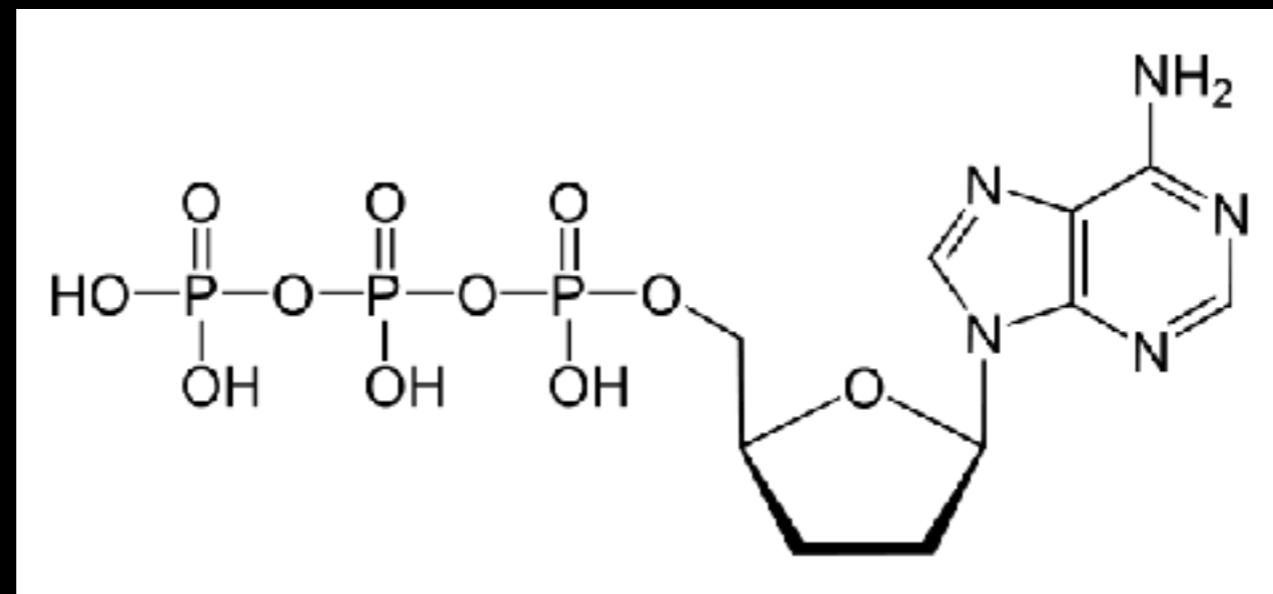
Introduced in 1977

Enzymatic method (F. Sanger, et al 1977) – automated sequencing

DNA POLYMERIZATION REACTION



ddDNA Sequencing



2',3'-dideoxyadenosine triphosphate (ddATP)¹

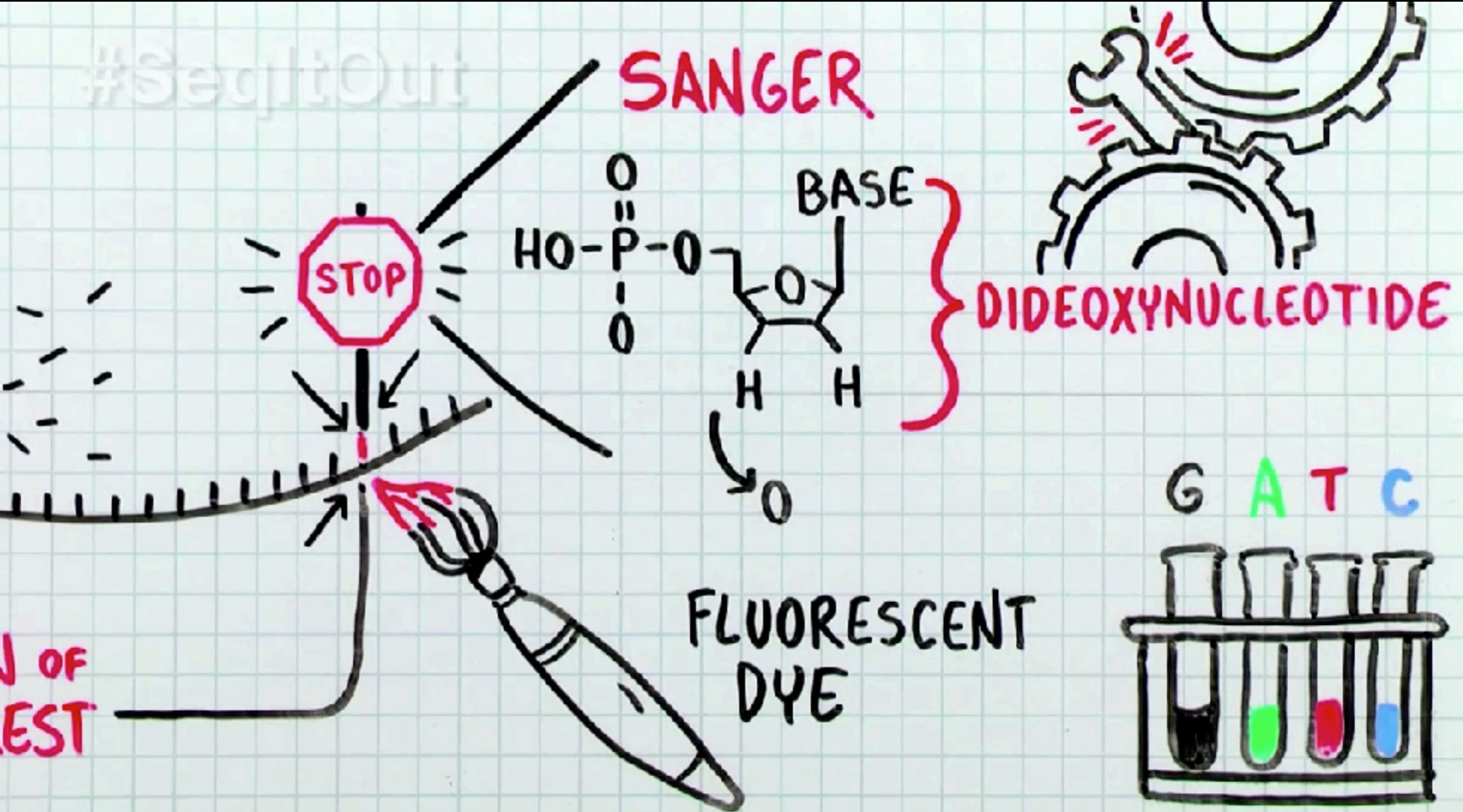
1) Wikipedia

Sanger Sequencing



<https://www.thermofisher.com/blog/behindthebench/how-does-sanger-sequencing-work/>

Sanger Sequencing



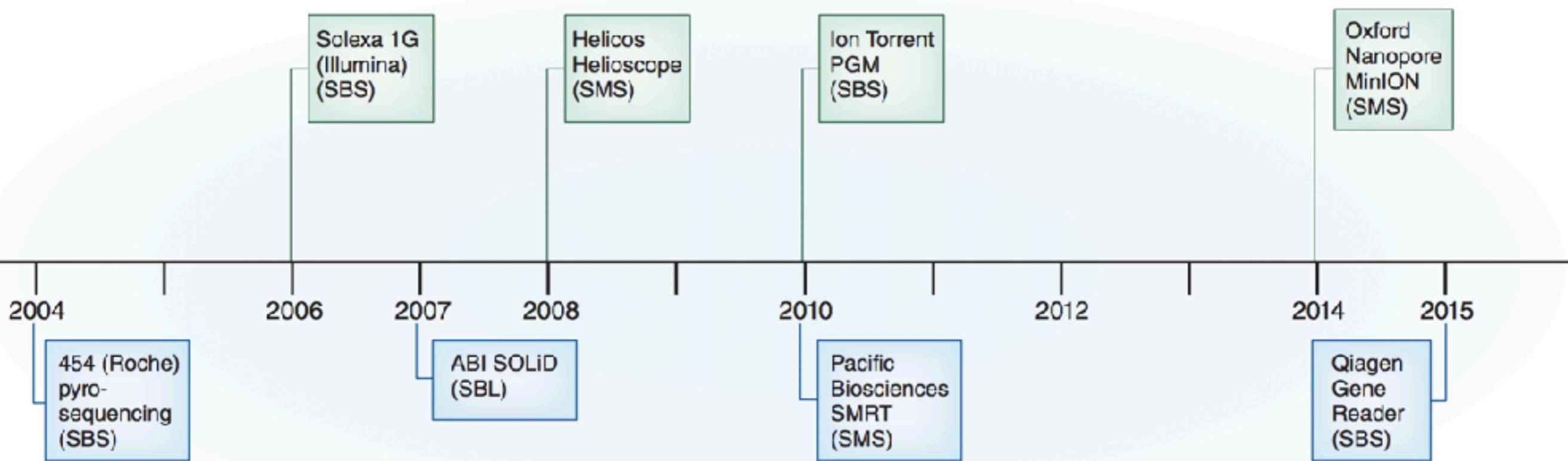
<https://www.thermofisher.com/blog/behindthebench/how-does-sanger-sequencing-work/>

Sanger Sequencing vs the next-generation sequencing

- Two decoupled steps vs integrated process
 - 1.produce the fluorescent signal
 - 2.data collection for analysis
- Sequence length

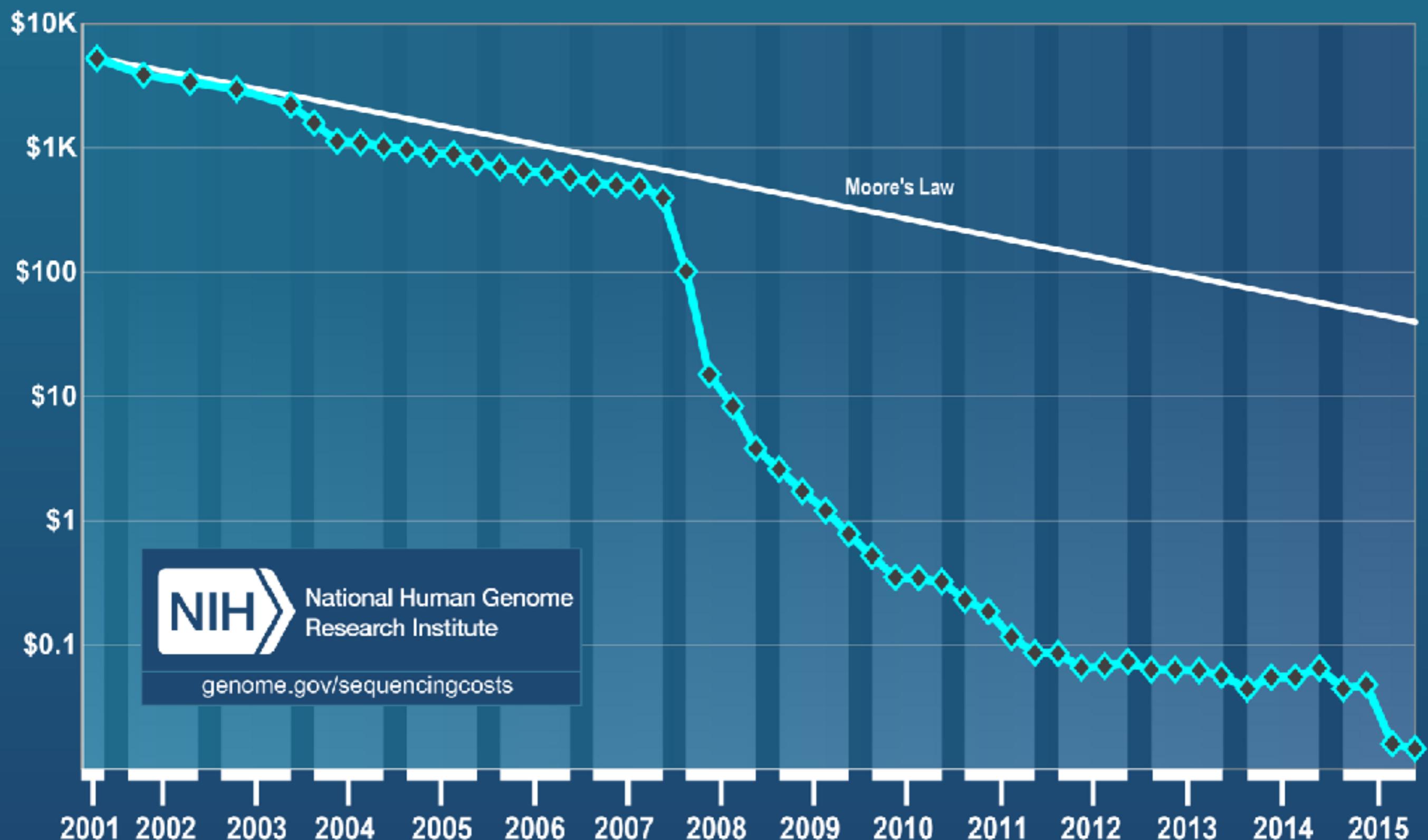
Next-Generation Sequencing

- High-throughput sequencing (as compared to the low throughput Sanger Sequencing).
- Massively parallel short-gun sequencing approach
 - Short-read NGS (35-700bp)
 - Sequencing by ligation (SBL)
 - Sequencing by synthesis (SBS)
 - Long-read NGS
 - lower throughput than short-read NGS



NGS instruments introduced over the past decade. This timeline describes the year of introduction of each of the NGS platforms that successfully achieved commercial introduction during the past decade. SBS, sequencing by synthesis; SMS, single-molecule sequencing; SBL, sequencing by ligation.

Cost per Raw Megabase of DNA Sequence

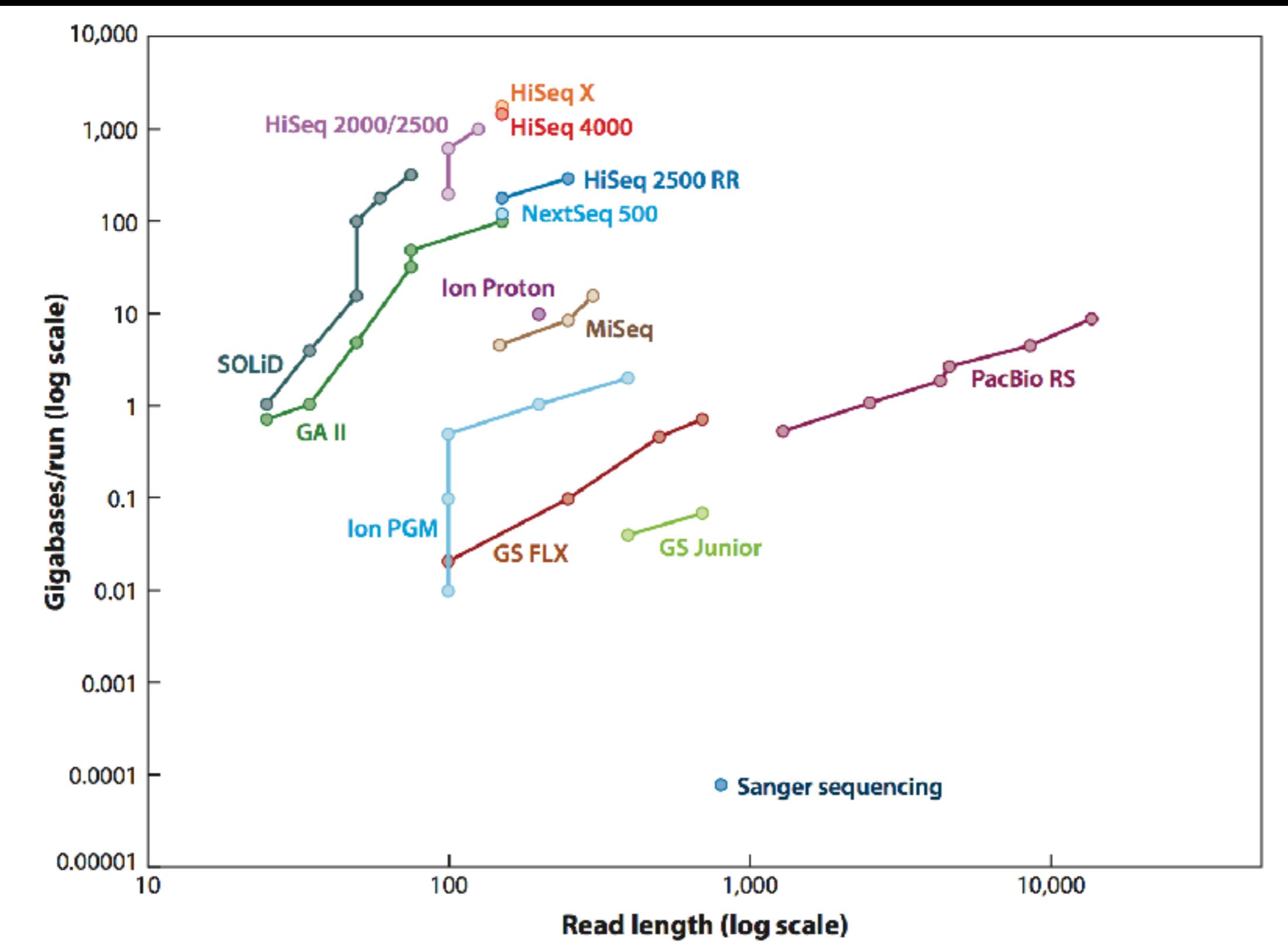


Throughput & Read-Length

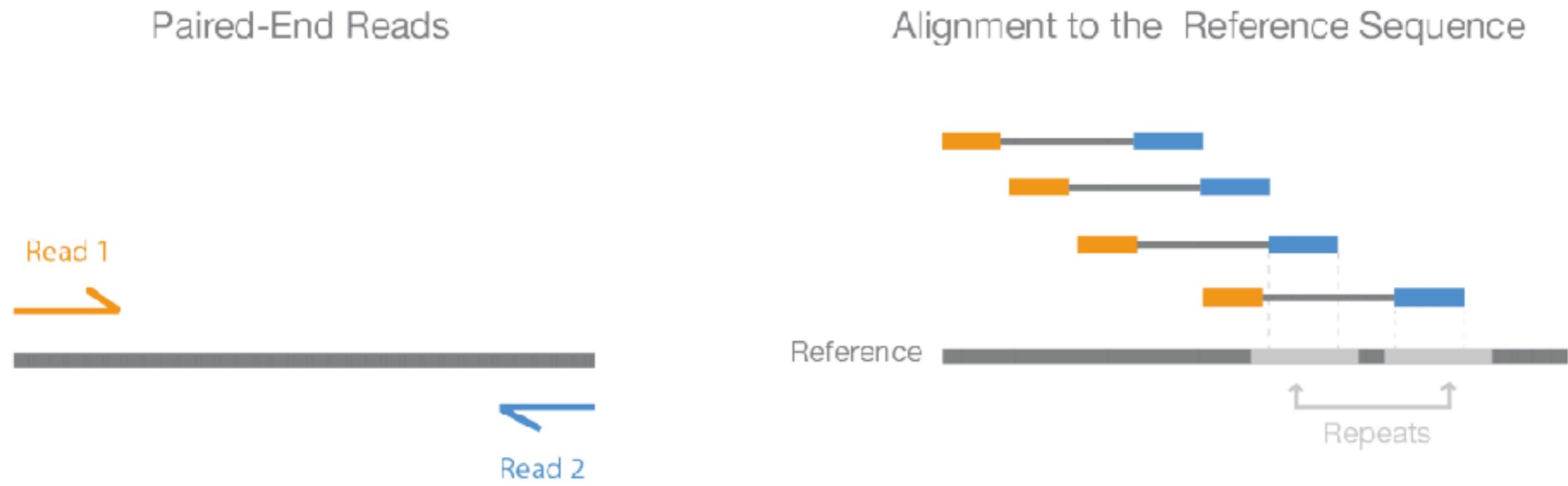
Max Throughput = Number of reads x Read-Length

- Throughput = Total throughput of a sequencing machine per run
- Read-Length = the length of the short-reads being generated by the machine

Read Length



Paired-end vs Single-end



Paired-end sequencing enables both ends of the DNA fragment to be sequenced, hence, facilitates detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts.

Common Steps for NGS

1. Create clonal template

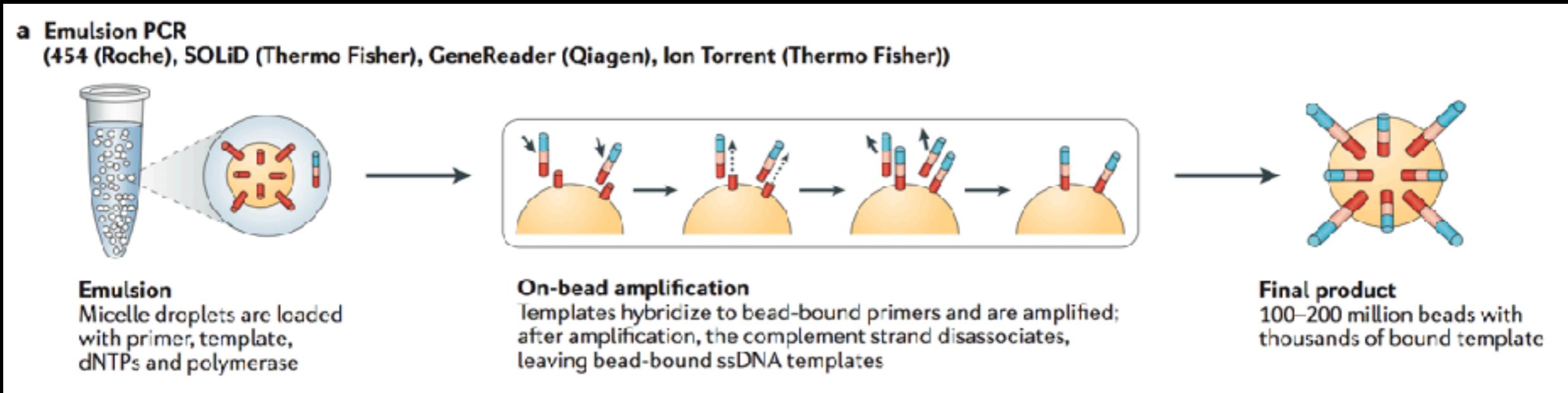
A. Fragmentation

B. Ligation to a common adapter

2. Template amplification

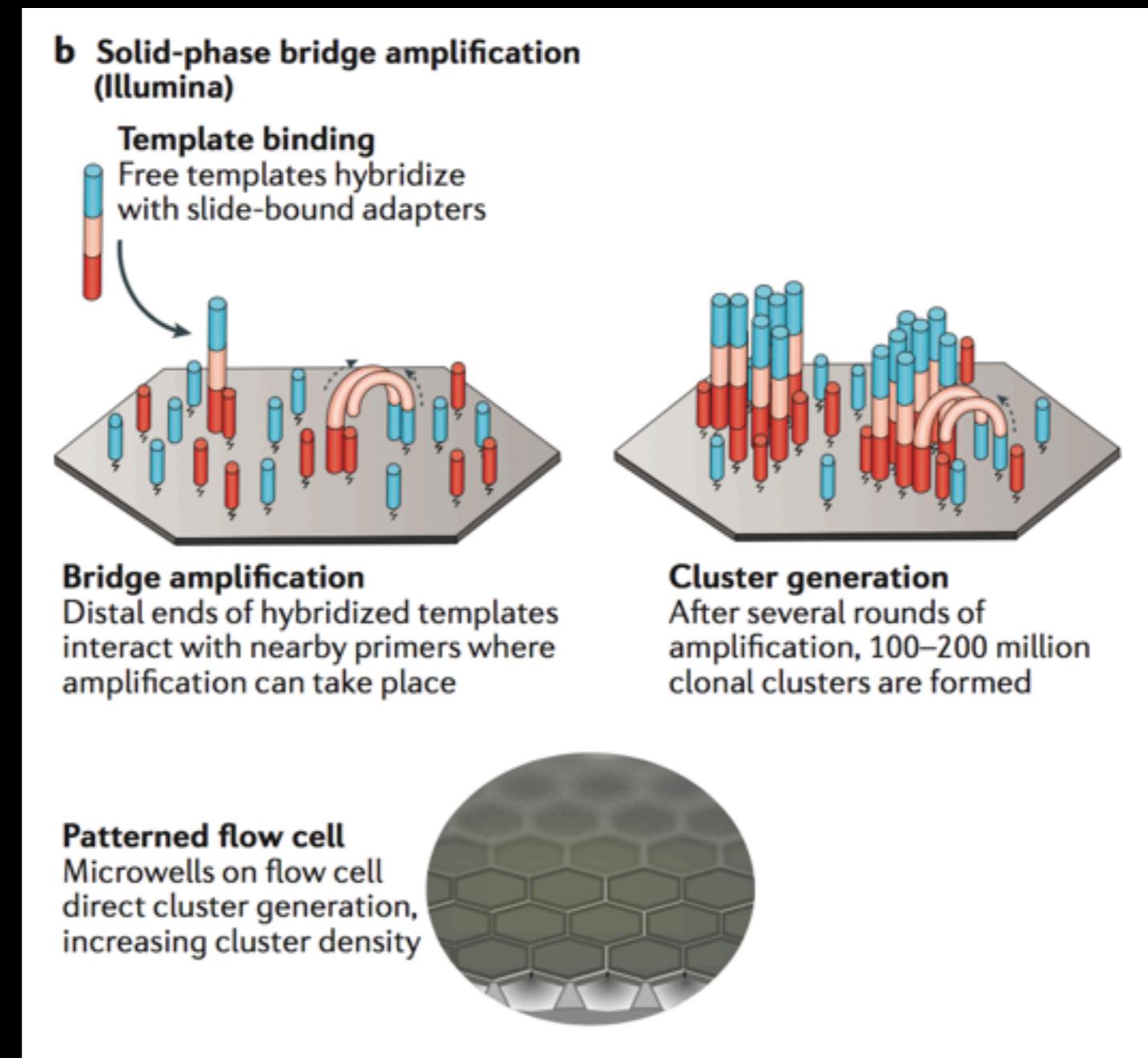
3. Massively Parallel Sequencing

Emulsion PCR



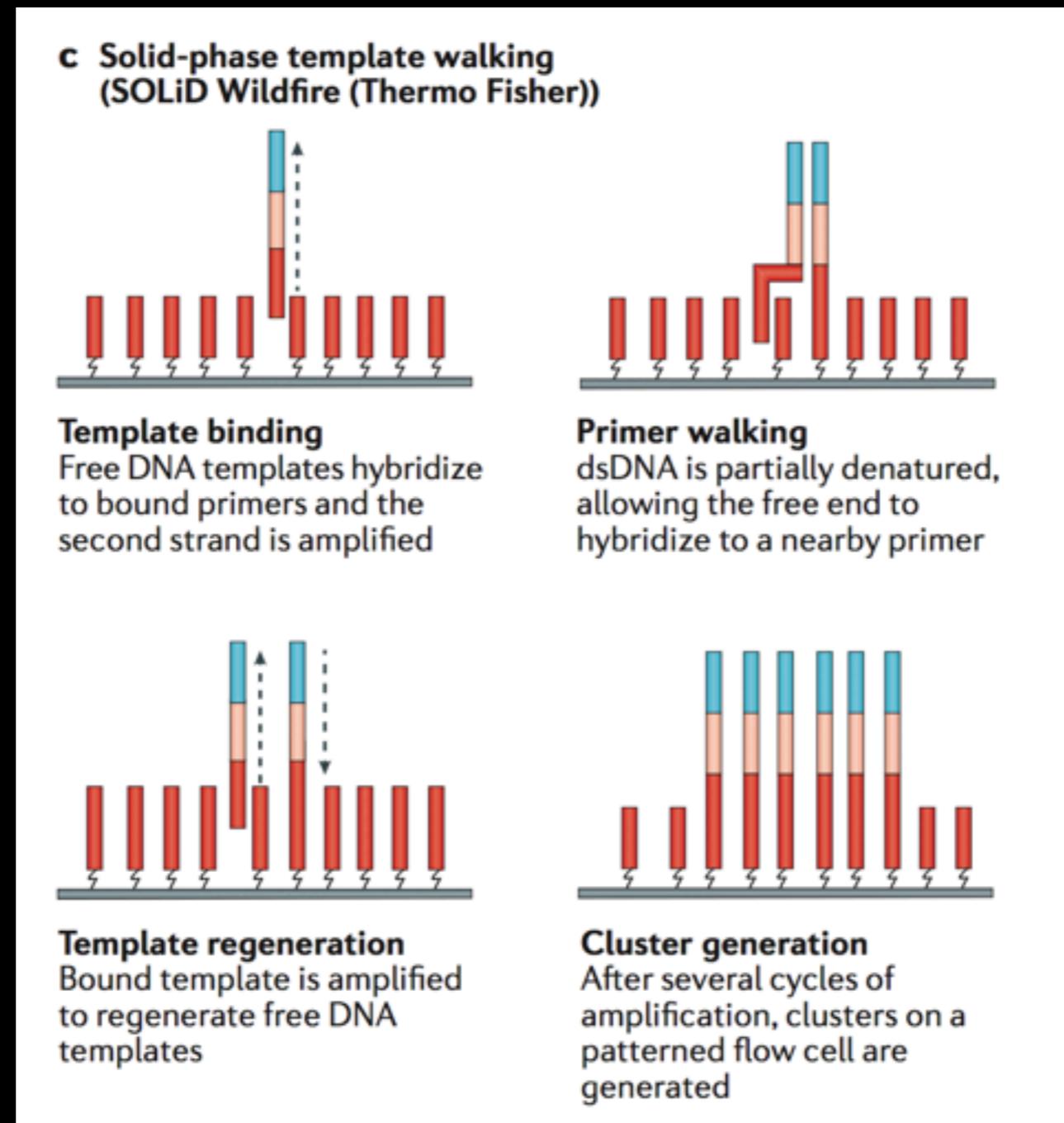
454 (Roche), SoLiD (Thermo Fisher),
GeneReader (Qiagen), Ion Torrent (Thermo Fisher)

Solid-phase bridge amplification



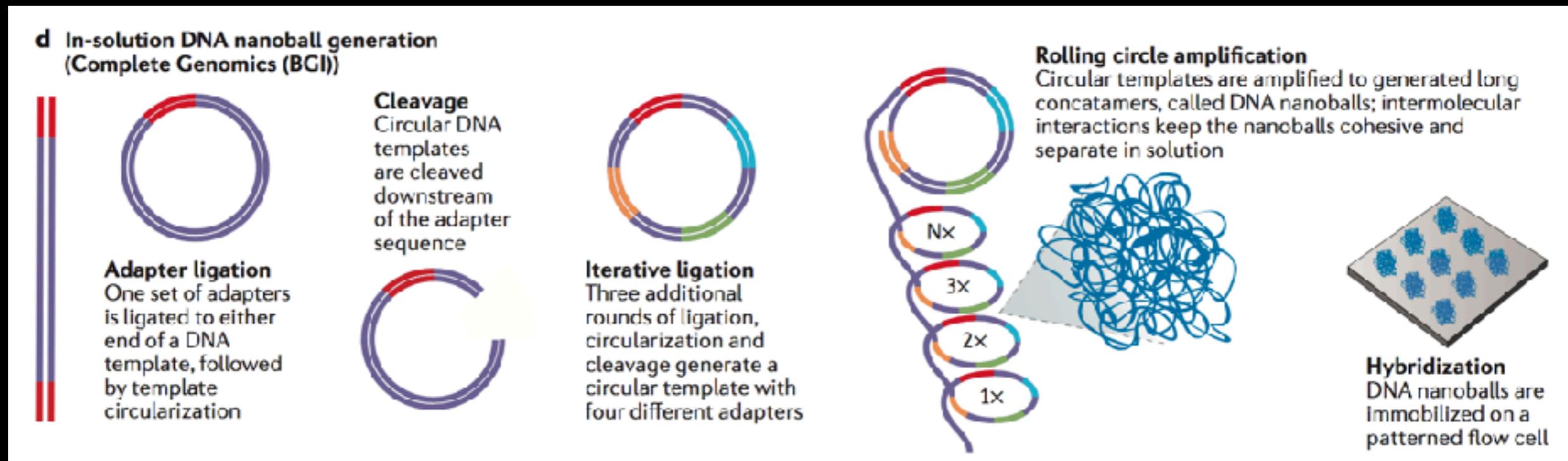
Illumina

Solid-phase template walking



SOLiD Wildfire

DNA nanoball amplification



BGI's Complete Genomics Platform

Hybridization Target Enrichment

- NimbleGen SeqCap EZ Choice
- NimbleGen Array capture method
- Roche SeqCap® target enrichment products
- Agilent SureSelect® target enrichment
- Custom regions
 - Region-specific enrichment — 20mer probes with Streptavidin bead (HMC region: Daprich, 2016), comparison (Garcia-Garcia, 2016)

Sequencing by Synthesis

Cyclic reversible termination
vs
Single-nucleotide addition

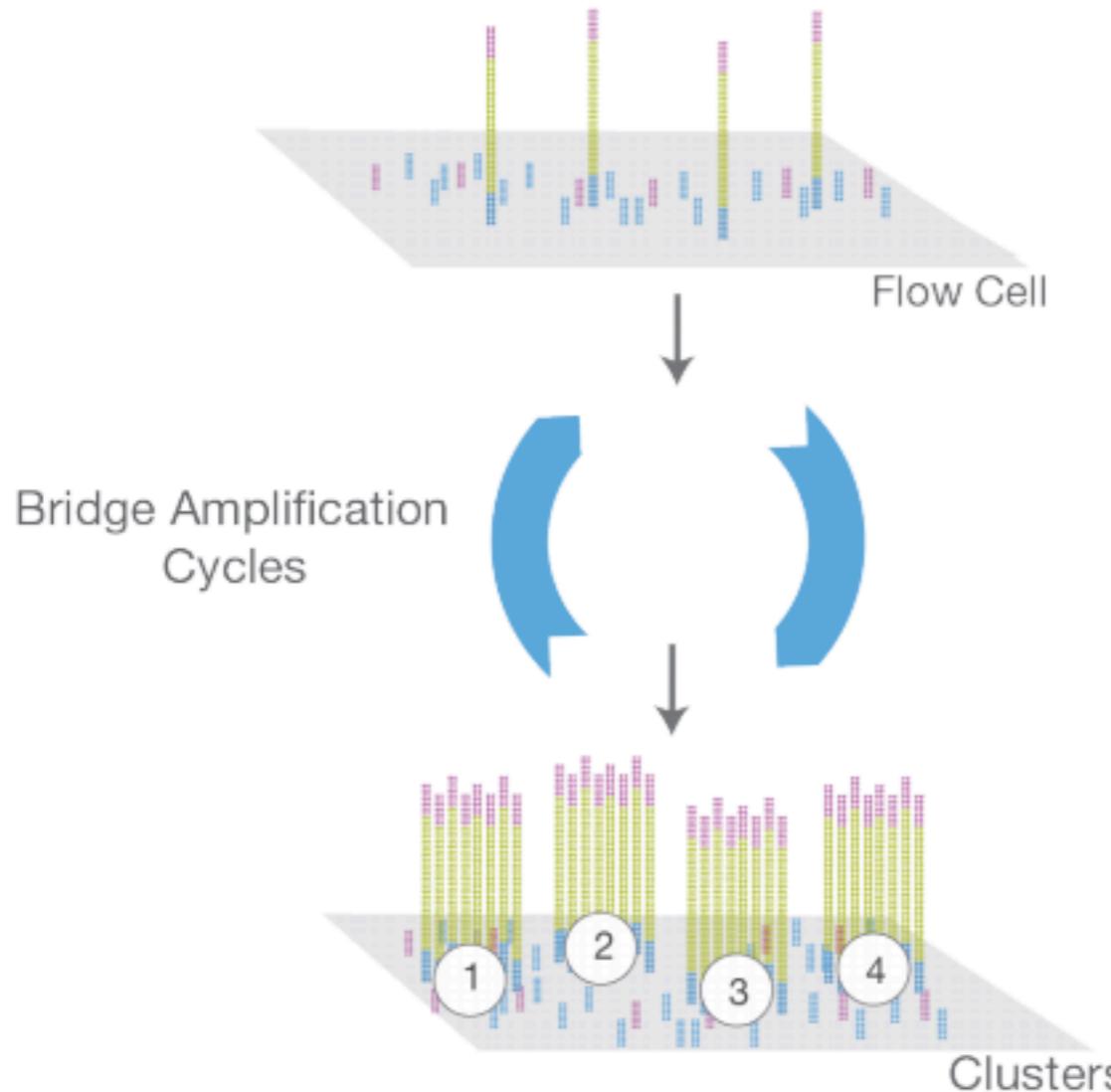
Illumina, Qiagen
vs
454, Ion Torrent

Illumina: HiSeq, MiSeq, NextSeq, MiniSeq

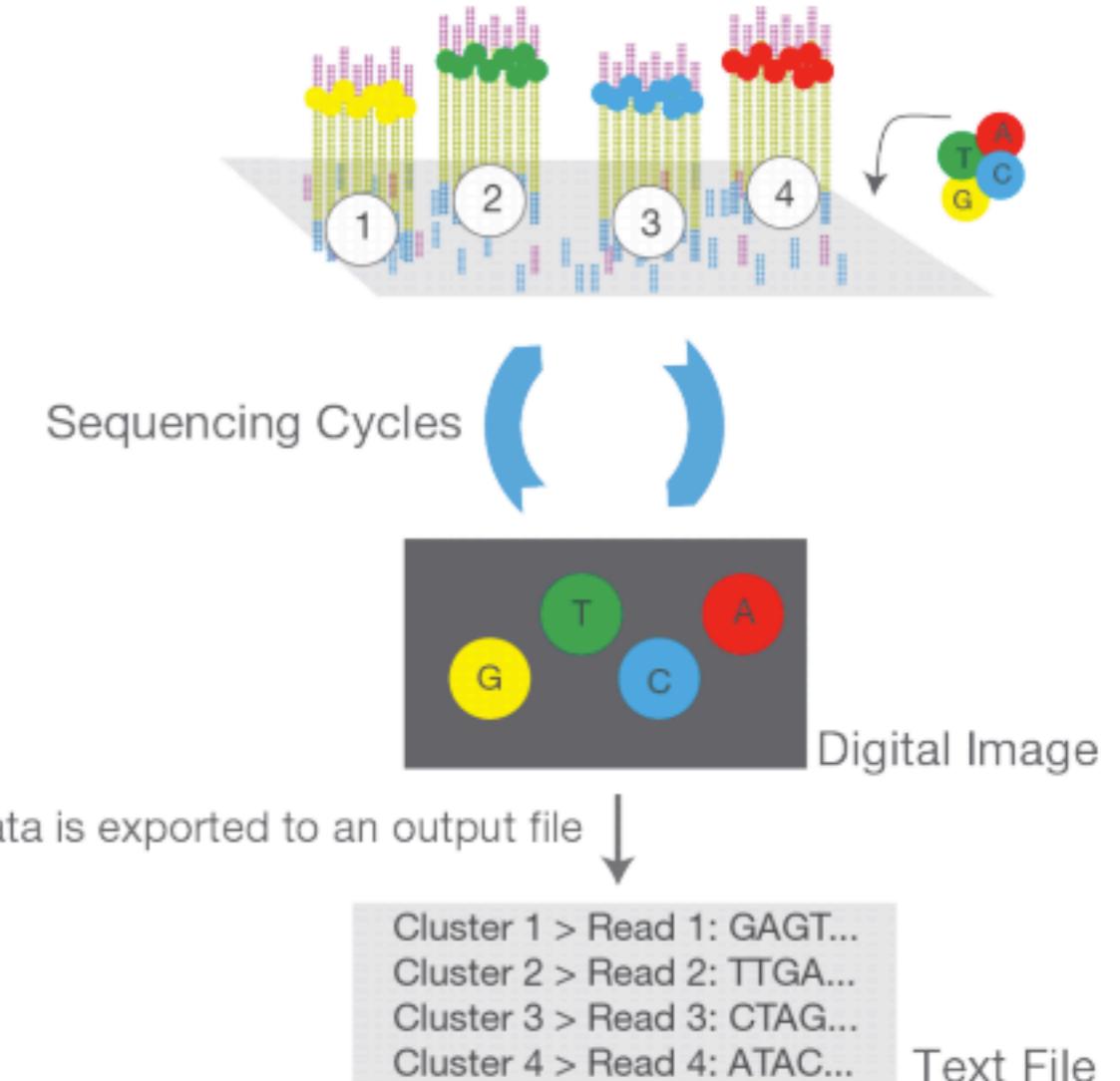
Qiagen: Intelligent BioSystems —> GeneReader

Sequencing by Synthesis

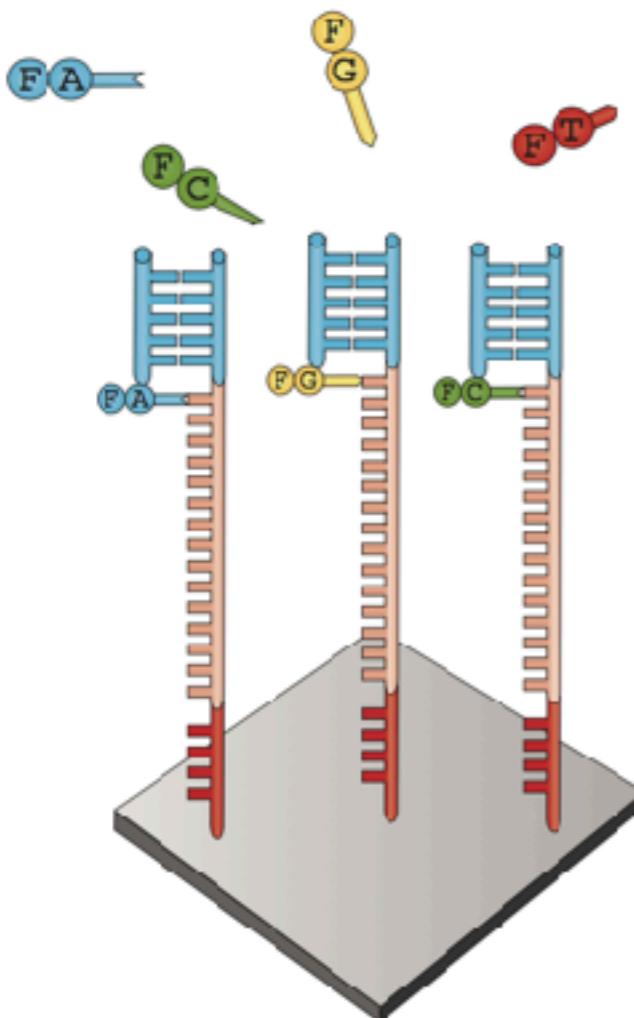
B. Cluster Amplification



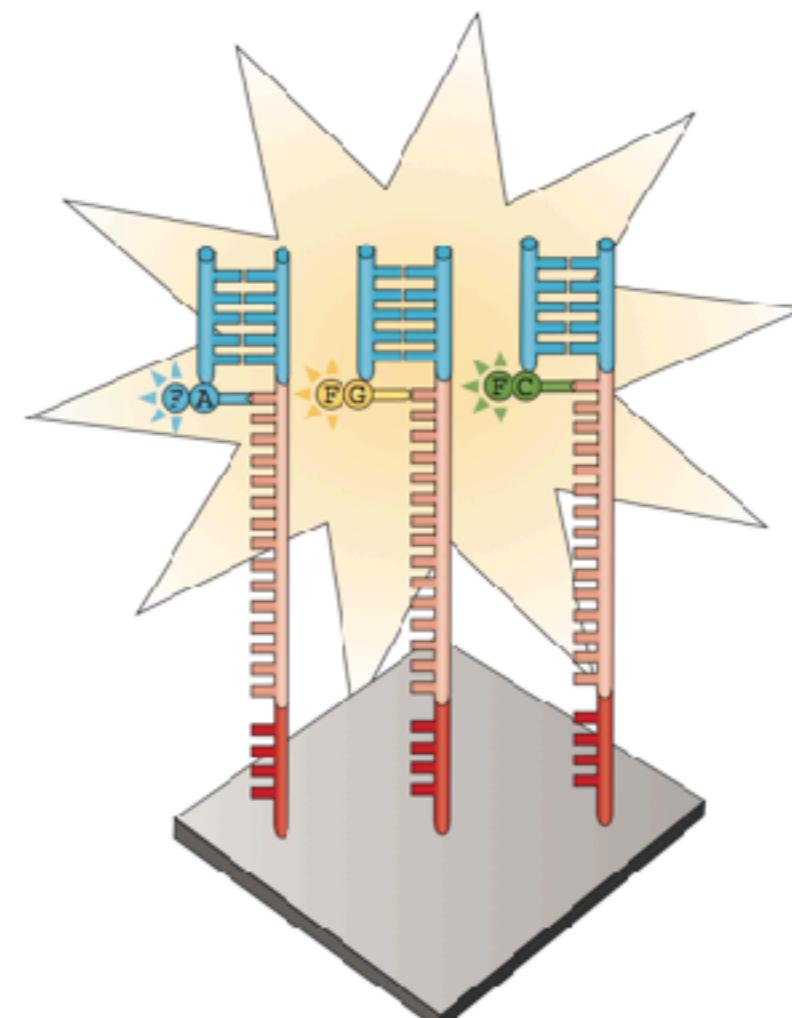
C. Sequencing



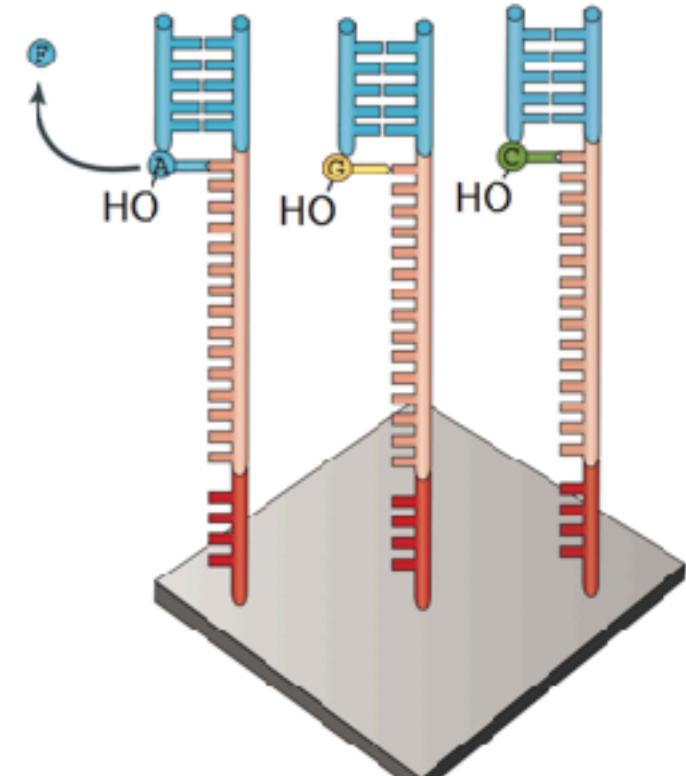
Sequencing by Synthesis

a Illumina**Nucleotide addition**

Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.

**Imaging**

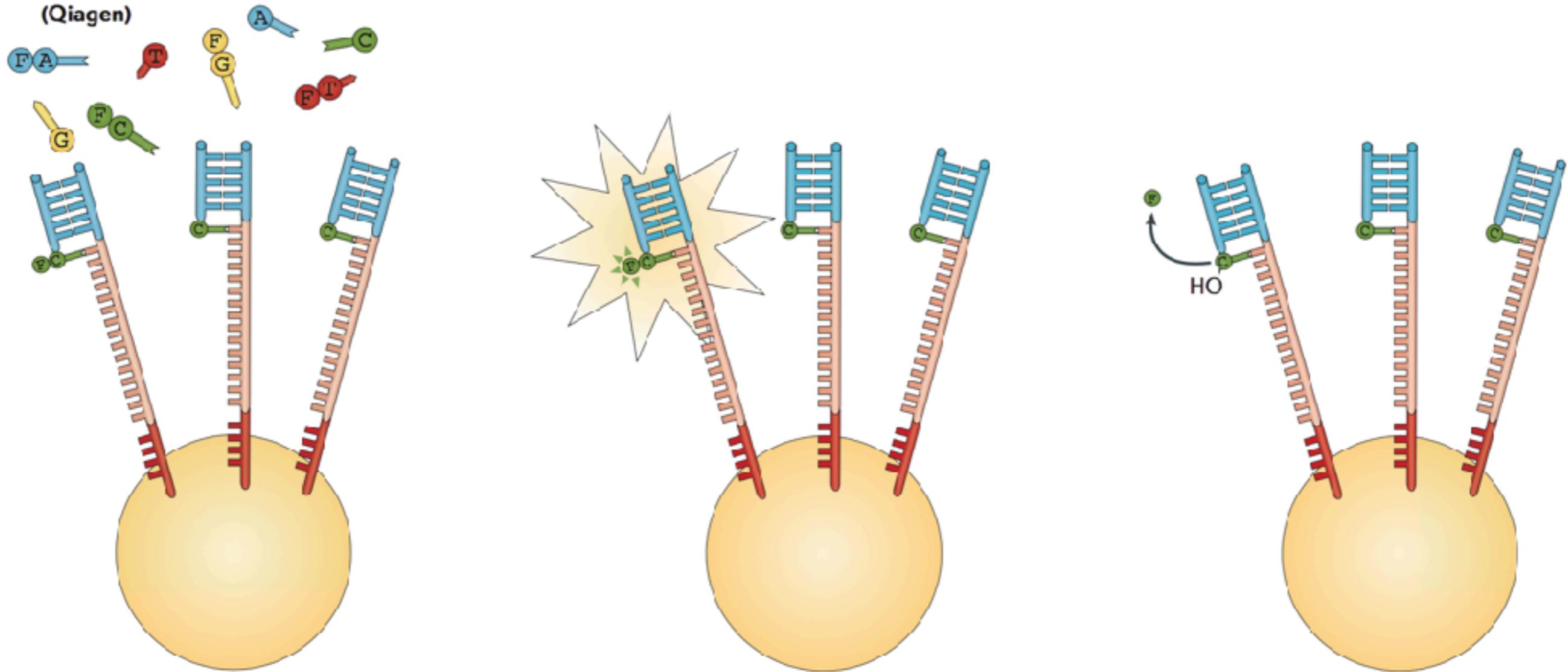
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.

**Cleavage**

Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Sequencing by Synthesis

b GeneReader
(Qiagen)



Nucleotide addition

A mixture of fluorophore-labelled, terminally blocked nucleotides and unlabelled, blocked nucleotides hybridize to complementary bases. Each bead on a slide can incorporate a different base.

Imaging

Slides are imaged with four laser channels. Each bead emits a colour corresponding to the base incorporated during this cycle, but only labelled bases emit a signal.

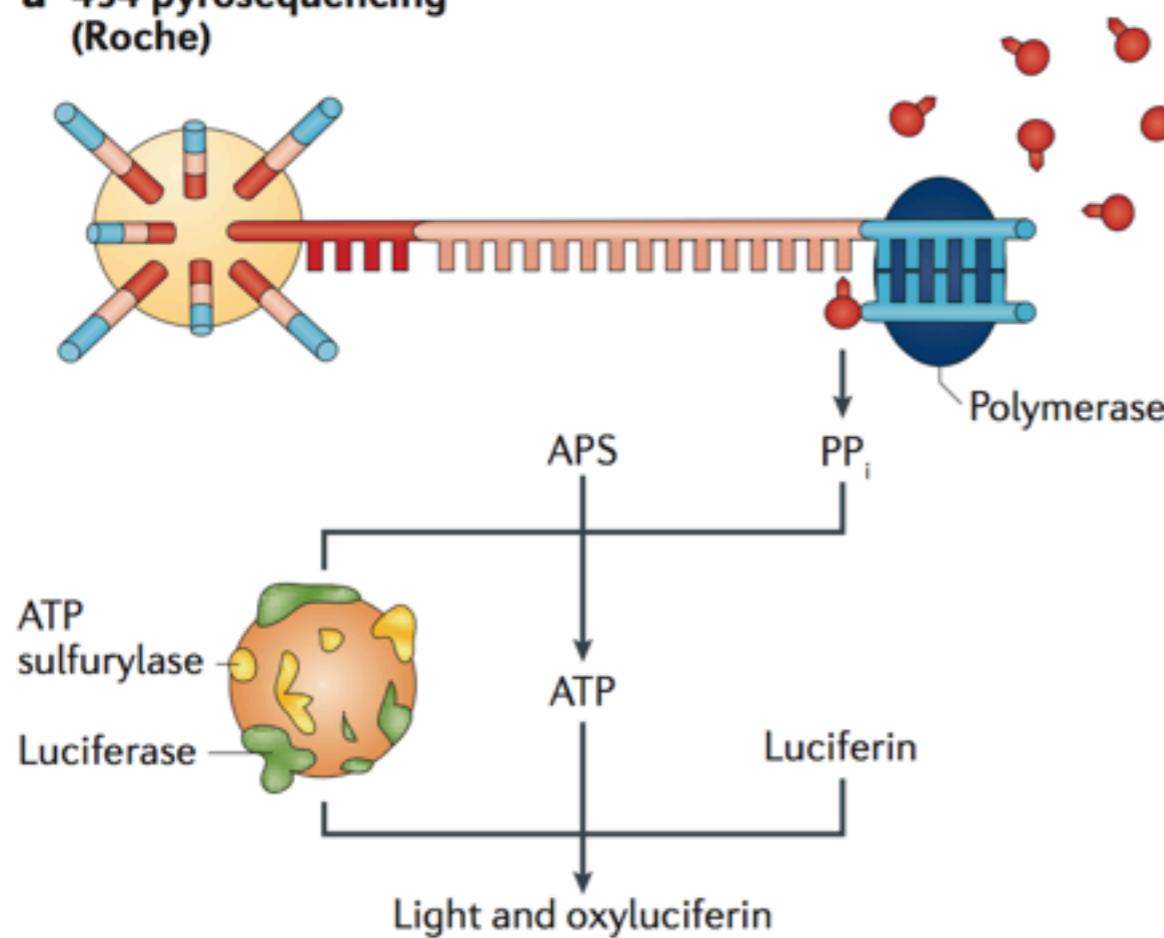
Cleavage

Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

454 pyrosequencing

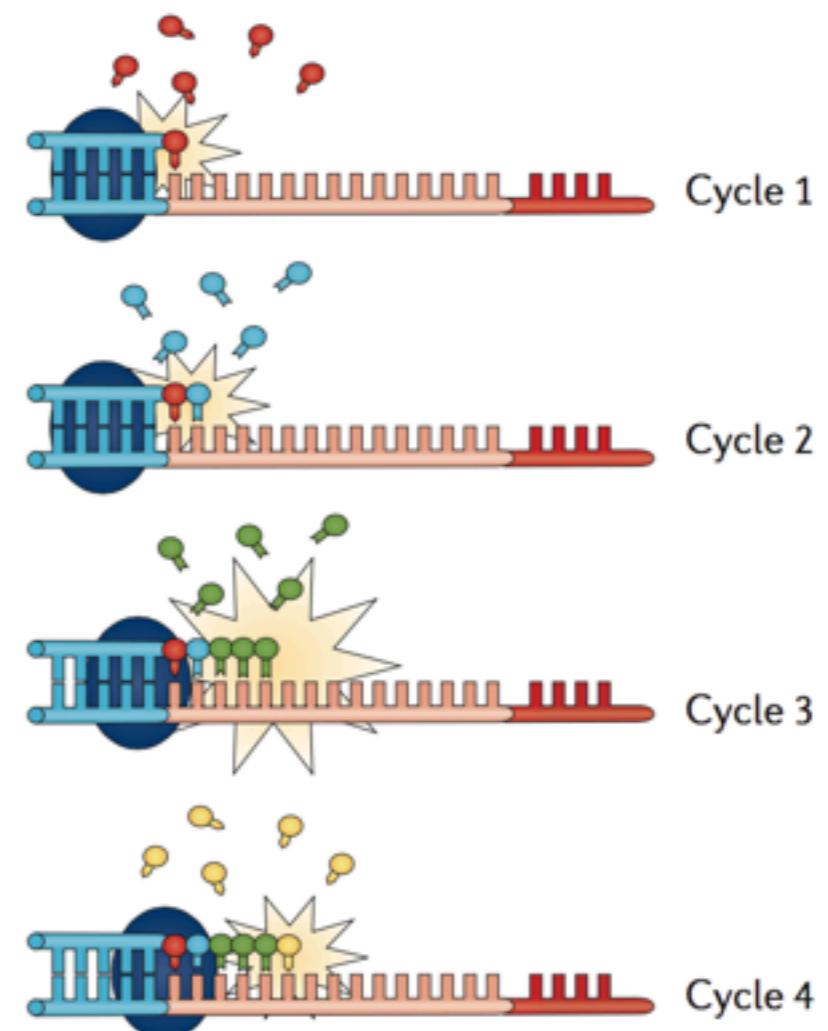
Sequencing by Synthesis

a 454 pyrosequencing (Roche)



Single nucleotide addition

Only one dNTP species is present during each cycle; multiple identical dNTPs can be incorporated during a cycle, increasing emitted light



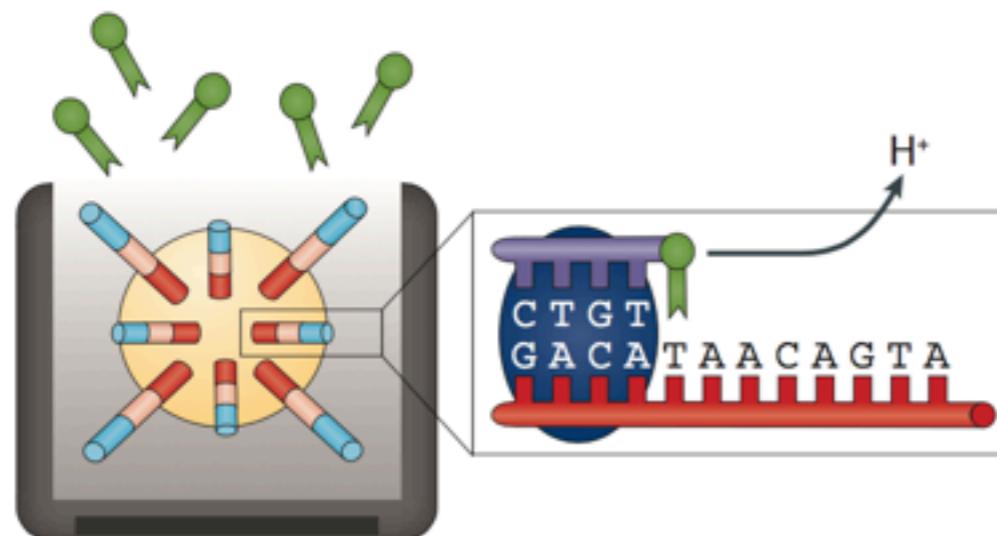
Pyrosequencing

As a base is incorporated, the release of an inorganic pyrophosphate triggers an enzyme cascade, resulting in light

Limited accuracy in detecting homopolymers

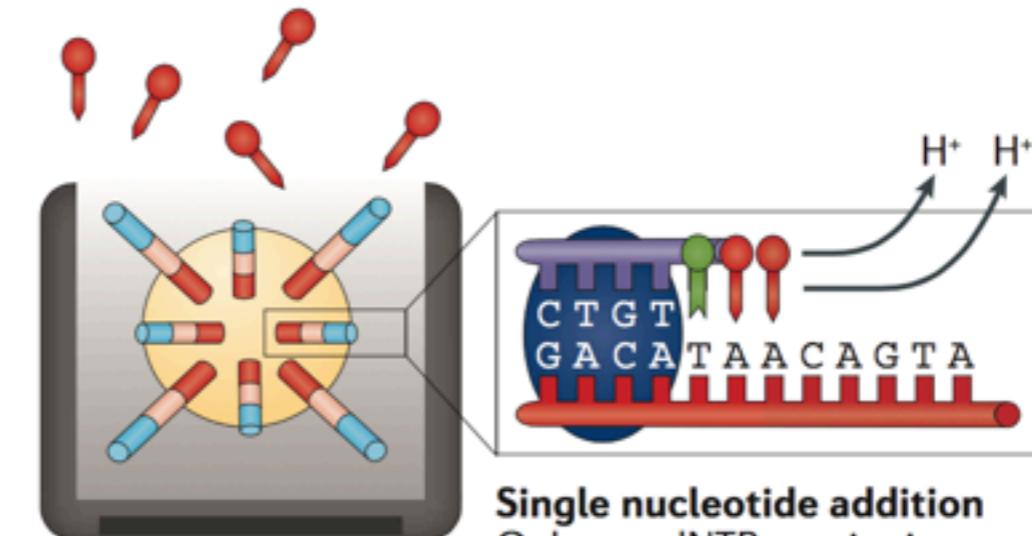
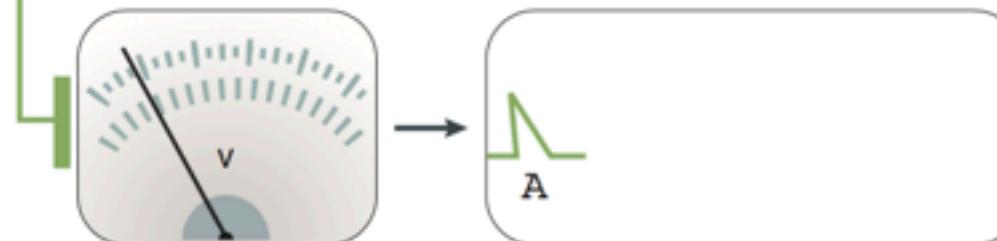
Sequencing by Synthesis

b Ion Torrent
(Thermo Fisher)



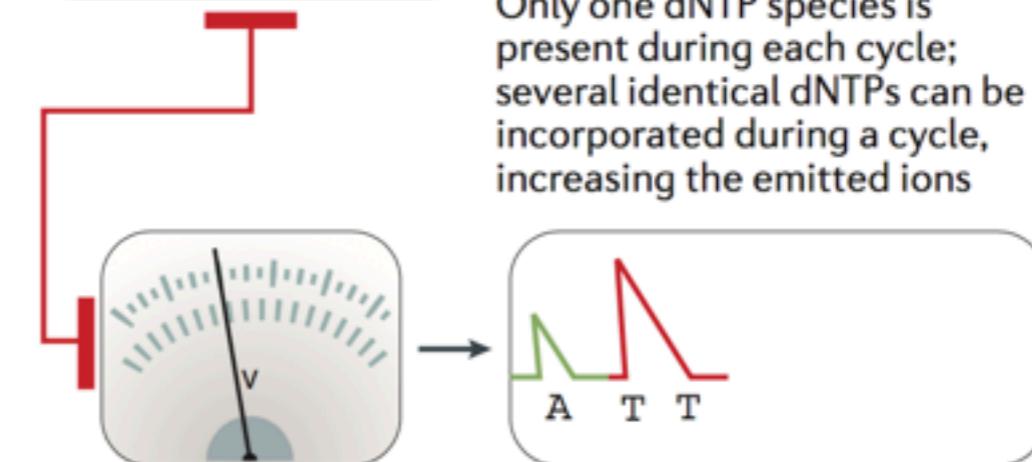
Semiconductor sequencing

As a base is incorporated, a single H^+ ion is released, which is detected by a CMOS-ISFET sensor



Single nucleotide addition

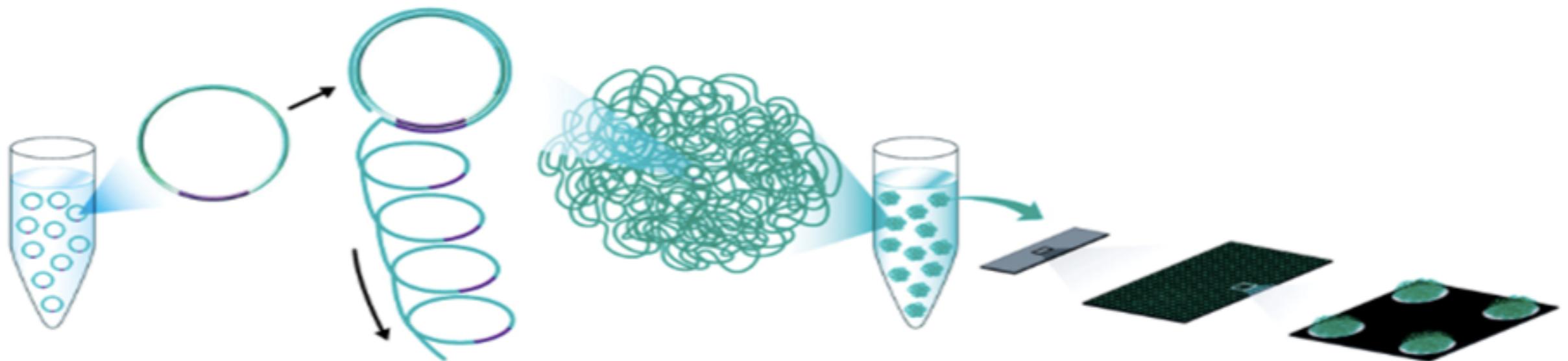
Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions



an integrated *Complementary Metal-Oxide-Semiconductor (CMOS)* and an *ion-sensitive field-effect transistor (ISFET)*

Limited accuracy in detecting homopolymers

Sequencing by Synthesis

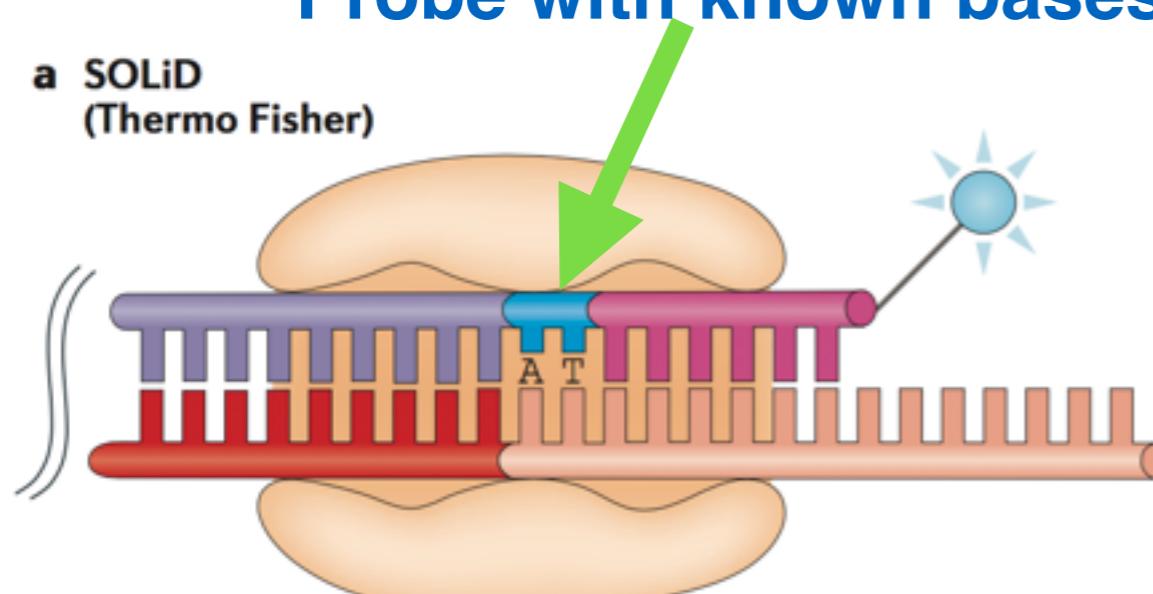


https://en.wikipedia.org/wiki/DNA_nanoball_sequencing

Sequencing by ligation

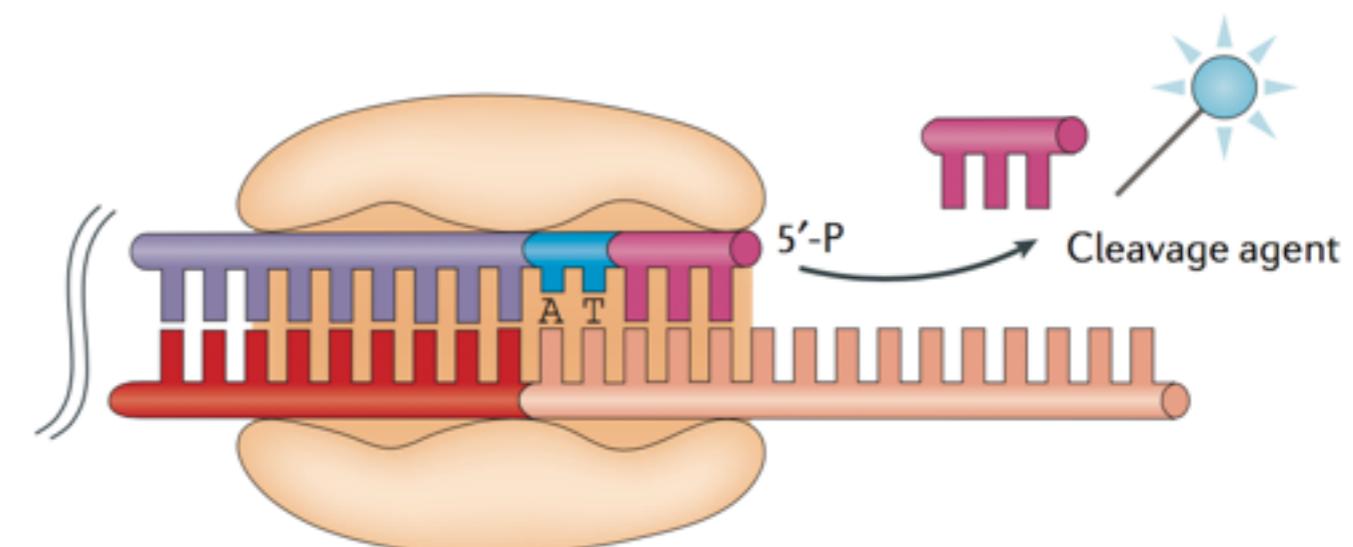
Probe with known bases

a SOLiD
(Thermo Fisher)



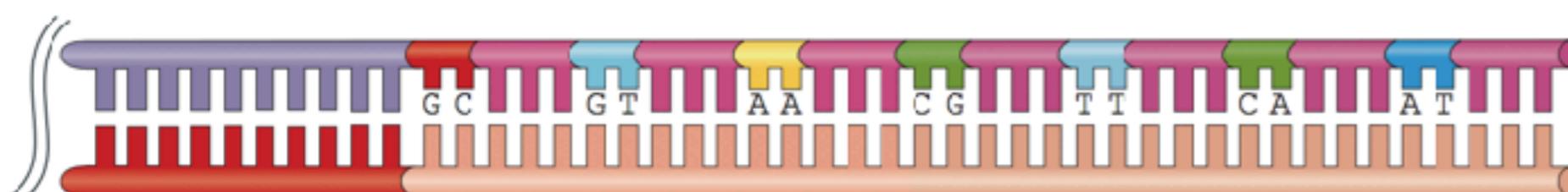
Two-base-encoded probes

Probes with two known bases followed by degenerate or universal bases hybridize to a template; ligase immobilizes the complex and the slide is imaged



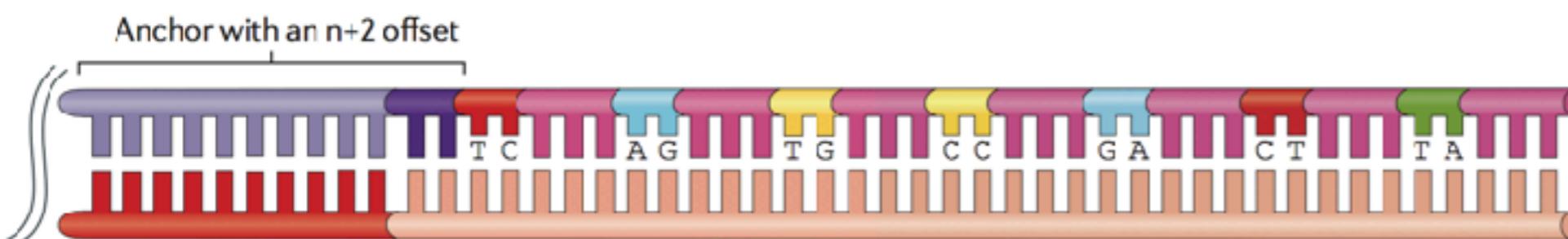
Cleavage

The fluorophore is cleaved from the probe along with several bases, revealing a 5' phosphate



Probe extension

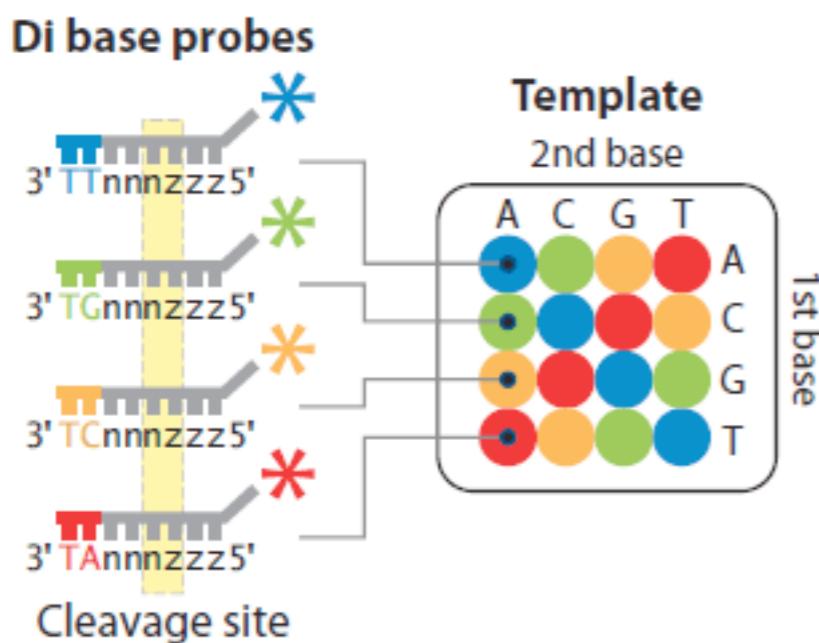
10 rounds of hybridization, ligation, imaging and cleavage identify 2 out of every 5 bases



Reset

After a round of probe extension, all probes and anchors are removed and the cycle begins again with an offset anchor

SOLID Color Space

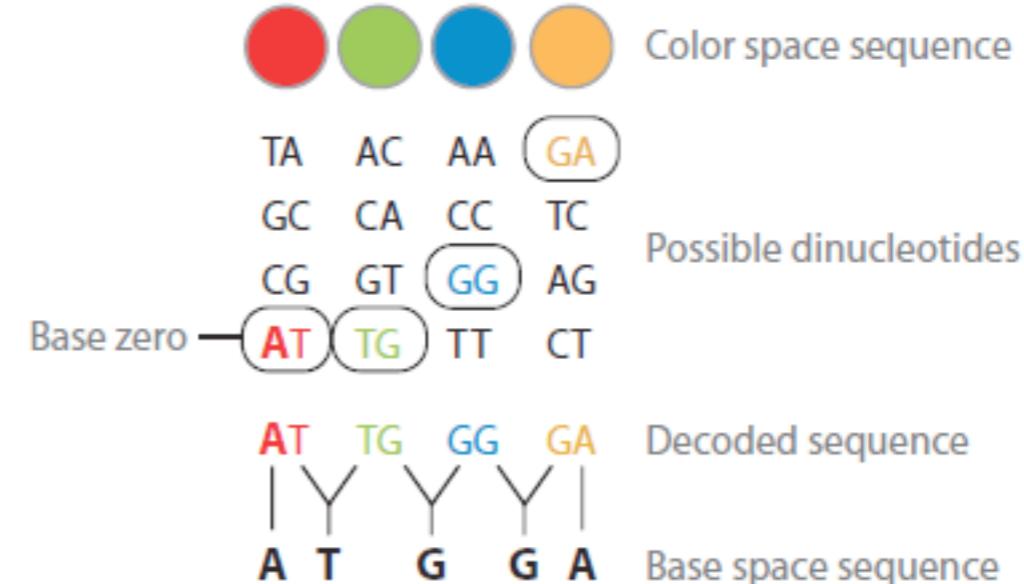


Double interrogation

With 2 base encoding each base is defined twice



Decoding



Long-read Sequencing

Single-molecule real-time sequencing (SMRT)
PacBio RS II & ONT MinIon/Prometheon

Synthetic long-read sequencing
Illumina (Moleculo) & 10X Genomics

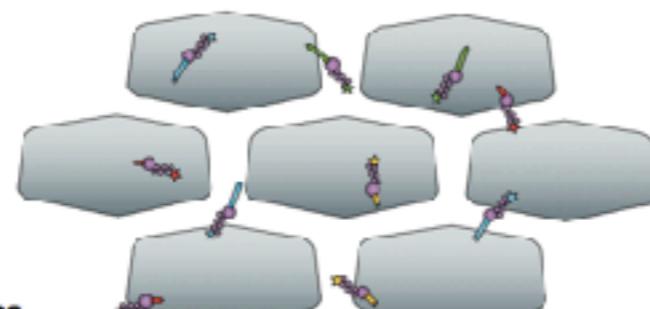
SMRT

Aa Pacific Biosciences

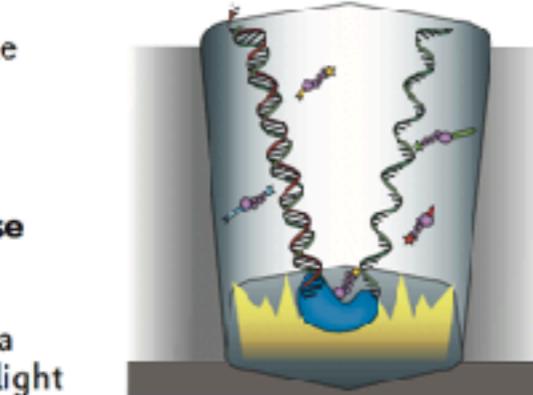
SMRTbell template
Two hairpin adapters allow continuous circular sequencing



ZMW wells
Sites where sequencing takes place



Labelled nucleotides
All four dNTPs are labelled and available for incorporation



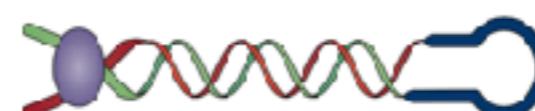
Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light

PacBio output
A camera records the changing colours from all ZMWs; each colour change corresponds to one base

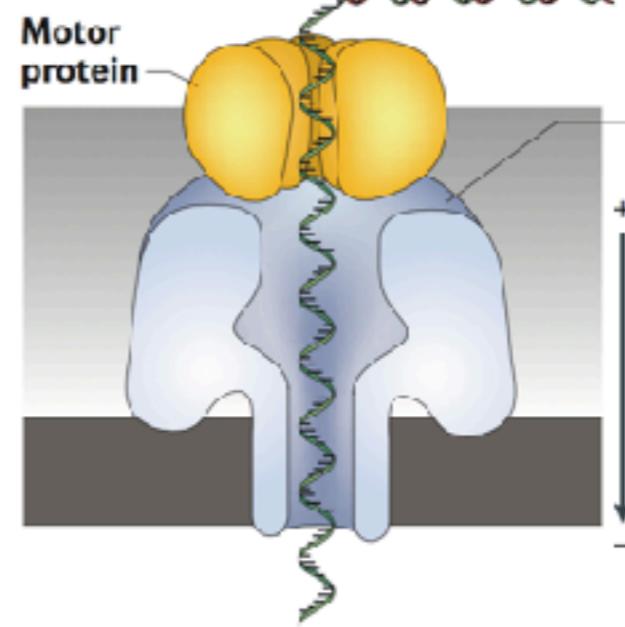


Ab Oxford Nanopore Technologies

Leader-Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing



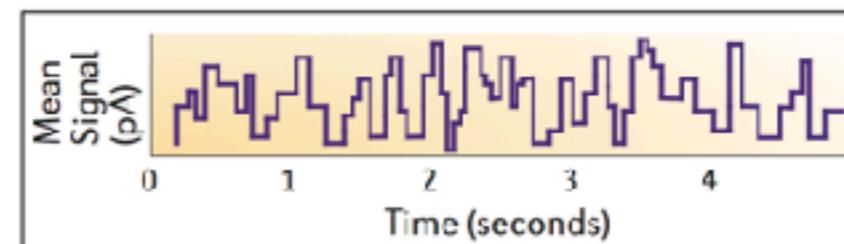
1D vs 2D strand



Motor protein

Alpha-hemolysin
A large biological pore capable of sensing DNA

Current
Passes through the pore and is modulated as DNA passes through



ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

squiggle space

Synthetic Long-read

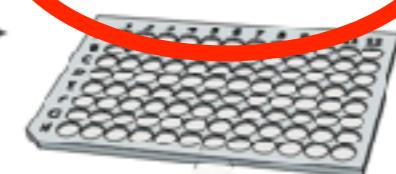
Ba Illumina

DNA fragment
DNA is fragmented and selected to ~10 kb

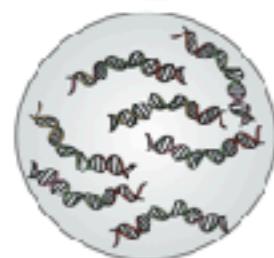


→
~3,000 molecules per well

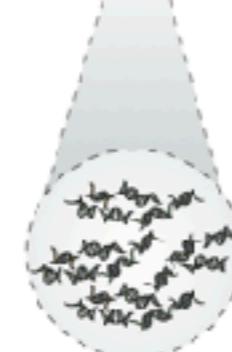
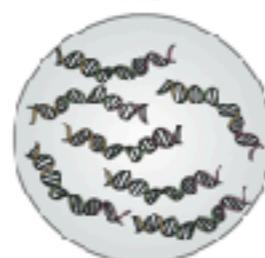
Enzymatic cleavage
DNA is barcoded and fragmented to ~350 bp



A1



A2



Barcodes

DNA from the same well shares the same barcode

Pooling
DNA from each well is pooled and undergoes a standard library preparation



Sequencing
DNA is sequenced on a standard short-read sequencer

One DNA sample per plate

1. Fragment DNA (~10kb)
2. Enzymatic cleavage (~350bp)

Pooling

3. Pooling multiple wells with unique well barcode

Assembly the same barcode into long-read

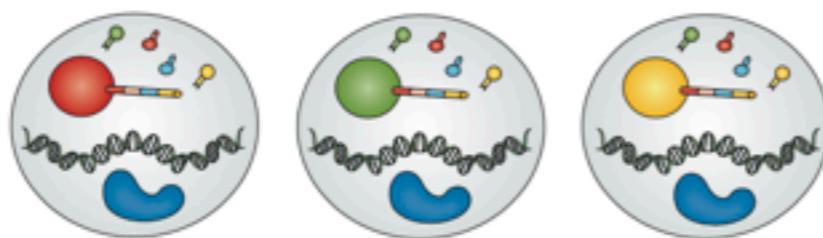
Assemble by the same barcode into long-read

Synthetic Long-read

Bb 10X Genomics

Emulsion PCR

Arbitrarily long DNA is mixed with beads loaded with barcoded primers, enzyme and dNTPs



Required GemCode & Chromium

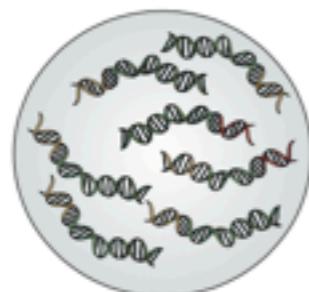
GEMs

Each micelle has 1 barcode out of 750,000



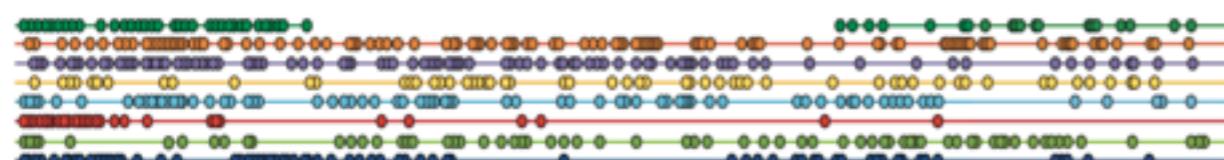
Amplification

Long fragments are amplified such that the product is a barcoded fragment ~350 bp



Pooling

The emulsion is broken and DNA is pooled, then it undergoes a standard library preparation



Assemble by common barcode linked reads

Single cell DNA sample

1. Emulsion PCR is done 1 cell/micelle, 1 barcode / micelle

Pooling

2. Pooling multiple micelles each with unique well barcode for sequencing

Assembly the same barcode into long-read

Illumina vs Qiagen GeneReader

Table 1 | Summary of NGS platforms

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
Sequencing by synthesis: CRT							
Illumina MiSeq Mid output	150 (SE)*	2.1–2.4 Gb*	14–16 M*	17 h*	<1%, substitution†	\$50,000 (REF. 118)	\$200–300 (REF. 118)
Illumina MiSeq High output	75 (SE)	1.6–1.8 Gb	22–25 M (SE)*	7 h	<1%, substitution†	\$50,000 (REF. 118)	\$200–300 (REF. 118)
	75 (PE)	3.3–3.7 Gb	44–50 M (PE)*	13 h			
	150 (PE)*	6.6–7.5 Gb*		24 h*			
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M (SE)	4 h	0.1%, substitution†	\$99,000‡	~\$1,000
	25 (PE)	750–850 Mb	24–30 M (PE)*	5.5 h			\$996
	150 (PE)	4.5–5.1 Gb		24 h			\$212
	250 (PE)*	7.5–8.5 Gb*		39 h*			\$142‡
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb	44–50 M (PE)*	21–56 h*	0.1%, substitution†	\$99,000‡	\$250
	300 (PE)*	13.2–15 Gb*					\$110‡
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb	Up to 260 M (PE)*	15 h	<1%, substitution†	\$250‡	\$42
	150 (PE)*	32–40 Gb*		26 h*			\$40‡
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb	400 M (SE)*	11 h	<1%, substitution†	\$250‡	\$43
	75 (PE)	50–60 Gb	800 M (PE)*	18 h			\$41
	150 (PE)*	100–120 Gb*		29 h*			\$33‡
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11 Gb	300 M (SE)*	7 h	0.1%, substitution‡	\$690‡	\$230
	50 (PE)	25–30 Gb	600 M (PE)*	16 h			\$90
	100 (PE)	50–60 Gb		27 h			\$52
	150 (PE)	75–90 Gb		40 h			\$45
	250 (PE)*	125–150 Gb*		60 h‡			\$40‡
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution‡	\$690‡	\$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			\$78
	100 (PE)*	270–300 Gb		11 d*			\$45‡
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution‡	\$690‡	\$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			\$58
	100 (PE)	360–400 Gb		5 d			\$45
	125 (PE)*	450–500 Gb*		6 d*			\$30‡
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution‡	\$740/\$900 (REF. 156)	\$50
	75 (PE)	325–375 Gb					\$31
	150 (PE)*	650–750 Gb*					\$22 (REF. 157)
Illumina HiSeq X	150 (PE)*	800–900 Gb per flow cell*	2.6–3 B (PE)*	<3 d*	0.1%, substitution‡	\$1,000‡	\$7.0‡
Qiagen GeneReader	NA‡	12 genes; 1,250 mutations‡‡	NA‡	Several days‡‡	Similar to other SBS systems‡‡	NA‡	\$400–\$600 per panel‡‡

SBL & SNA

Table 1 | Summary of NGS platforms

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
<i>Sequencing by ligation</i>							
SOLID 5500 Wildfire	50 (SE)	80 Gb	~700M*	6 d*	≤0.1%, AT bias [#]	NA ⁱ	\$130 ^f
	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLID 5500xl	50 (SE)	160 Gb	~1.4B*	10 d*	≤0.1%, AT bias [#]	\$251,000 ^e	\$70 ^f
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					
BGISEQ-500 FCS ¹⁵⁴	50–100 (SE/PE)*	8–40Gb ^a	NA ^j	24 h*	≤0.1%, AT bias [#]	\$250 (REF 155)	NA ⁱ
BGISEQ-500 FCL ¹⁵⁵	50–100 (SE/PE)*	40–200Gb*	NA ^j	24 h*	≤0.1%, AT bias [#]	\$250,000 (REF 155)	NA ⁱ
<i>Sequencing by synthesis: SNA</i>							
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel ^f	NA ⁱ	\$40,000 ^f
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel ^f	\$108,000 ^e	\$19,500 ^f
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)*	450 Mb*	~1 M*	10 h*	1%, indel ^f	NA ⁱ	\$15,500 ^f
454 GS FLX Titanium XL+	Up to 1,000; 700 mode (SE, PE)*	700 Mb*	~1 M*	23 h*	1%, indel ^f	\$450,000 ^e	\$9,500 ^f
Ion PGM 314	200 (SE)	30–50	400,000–550,000*	23 h	1%, indel ^f	\$49 ⁱ	\$25–3,500 ^f
	400 (SE)	60–100 Mb*		3.7 h*			
Ion PGM 316	200 (SE)	300–500 Mb	2–3 M*	3 h	1%, indel ^f	\$49 ⁱ	\$700–1,000 ^f
	400 (SE)*	600 Mb–1 Gb*		4.9 h*			
Ion PGM 318	200 (SE)	600 Mb–1 Gb	4–5.5 M*	4 h	1%, indel ^f	\$49 ⁱ	\$450–800 ^f
	400 (SE)*	1–2 Gb*		7.3 h*			
Ion Proton	Up to 200 (SE)	Up to 10 Gb*	60–80 M*	2–4 h*	1%, indel ^f	\$224 ^e	\$80 ^e
Ion S5 520	200 (SE)	600 Mb–1 Gb	3–5 M*	2.5 h	1%, indel ^f	\$65 (REF 158)	\$2,400 [*]
	400 (SE)*	1.2–2 Gb*		4 h*			\$1,200 [*]
Ion S5 530	200 (SE)	3–4 Gb	15–20 M*	2.5 h	1%, indel ^f	\$65 (REF 158)	\$950 [*]
	400 (SE)*	6–8 Gb*		4 h*			\$475 [*]
Ion S5 540	200 (SE)*	10–15 Gb*	60–80 M*	2.5 h*	1%, indel ^f	\$65 (REF 158)	\$300 [*]

**PGM Dx :
PE capability**

Single-molecule long reads

Table 1 | Summary of NGS platforms

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US\$)	Cost per Gb (US\$, approx.)
Single-molecule real-time long reads							
Pacific BioSciences RS II	~20 Kb	500 Mb–1 Gb*	~55,000*	4 h*	13% single pass, ≤1% circular consensus read, indel†	\$695‡	\$1,000‡
Pacific Biosciences Sequel	8–12 Kb ^{b9}	3.5–7 Gb*	~350,000*	0.5–6 h*	NA ^l	\$350 (REF. 69)	NA ^l
Oxford Nanopore MK 1 MinION	Up to 200 Kb ¹⁵⁹	Up to 1.5 Gb ¹⁵⁹	>100,000 (REF. 159)	Up to 48 h ¹⁶⁰	~12%, indel ¹⁵⁹	\$1,000*	\$750*
Oxford Nanopore PromethION	NA ^l	Up to 4 Tb*	NA ^l	NA ^l	NA ^l	\$75*	NA ^l
Synthetic long reads							
Illumina Synthetic Long-Read	~100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	No additional instrument required	~\$1,000*
10X Genomics	Up to 100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	\$75 (REFS 72, 151)	See HiSeq 2500 +\$500 per sample ¹⁵¹

Approx., approximate; AT, adenine and thymine; B, billion; bp, base pairs; d, days; Gb, gigabase pairs; h, hours; indel, insertions and deletions; Kb, kilobase pairs; M, million; Mb, megabase pairs; A, not available; , paired-end sequencing; B, sequencing by synthesis; , single-end sequencing; Tb, terabase pairs *Manufacturer's data. ‡Rounded from Field Guide to next-generation DNA sequencers¹⁶⁰ and 2014 update. §Not available as this instrument will be discontinued or only available as an upgraded version ||As this product has been developed only recently, this information is not available. ¶Not available as a single instrument.

Applications of NGS

- Whole genome sequencing (WGS)
 - 1000 Genomes Project
 - Population-level sequencing projects
- Whole-exome and targeted sequencing
 - Improve disease diagnosis
 - Discover new genes & diseases
- Additional insight to the genome
 - ***Assay for transposase-accessible chromatin using sequencing***
 - ATAC-seq, ChIP-seq, DNA methylation sequencing (methyl-seq)
- Single-cell omics

Platform & Applications of NGS

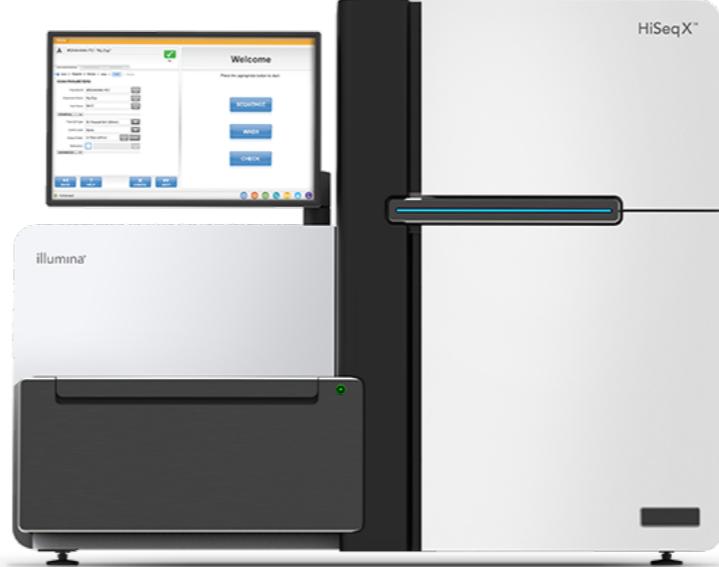
Company	Read length	Applications	Website
454/Roche	400 bp (single end)	Bacterial and viral genomes, multiplex-PCR products, validation of point mutations, targeted somatic-mutation detection	http://www.454.com/
Illumina	150–300 bp (paired end)	Complex genomes (human, mouse and plants) and genome-wide NGS applications, RNA-seq, hybrid capture or multiplex-PCR products, somatic-mutation detection, forensics, noninvasive prenatal testing	http://www.illumina.com/
ABI SOLiD	75 bp (single end) or 50 bp (paired end)	Complex genomes (human, mouse, plants) and genome-wide NGS applications, RNA-seq, hybrid capture or multiplex-PCR products, somatic-mutation detection	http://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html/
Pacific Biosciences	Up to 40 kb (single end or circular consensus)	Complex genomes (human, mouse and plants), microbiology and infectious-disease genomes, transcript-fusion detection, methylation detection	http://www.pacb.com/
Ion Torrent	200–400 bp (single end)	Multiplex-PCR products, microbiology and infectious diseases, somatic-mutation detection, validation of point mutations	http://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing.html/
Oxford Nanopore	Variable: depends on library preparation (1D or 2D reads)	Pathogen surveillance, targeted mutation detection, metagenomics, bacterial and viral genomes	http://nanoporetech.com/
Qiagen GeneReader	107 bp (single end)	Targeted mutation detection, liquid biopsy in cancer	http://www.genereaderngs.com/

Illumina Sequencing Machine



HiSeq (2500, 3000, 4000)

5B/1.5T



HiSeq X

6B/1.8T



NovaSeq 6000

20B/6T



iSeq 100

4M/1.2G



MiniSeq

25M/7.5G



MiSeq

25M/15G



NextSeq

400M/120G

Max Output/Max Reads (M = millions, G = Gigabyte, B = billions, T = Terabyte)

Illumina production-scale sequencer



Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)	●	●	●	●
Small Whole-Genome Sequencing (microbe, virus)	●	●		●
Exome Sequencing	●	●	●	●
Targeted Gene Sequencing (amplicon, gene panel)	●	●		●
Whole-Transcriptome Sequencing	●	●	●	●
Gene Expression Profiling with mRNA-Seq	●	●		●
miRNA & Small RNA Analysis	●	●		●
DNA-Protein Interaction Analysis	●	●		●
Methylation Sequencing	●	●		●
Shotgun Metagenomics	●	●		●

Thermo Scientific/Ion Torrent



Ion GeneStudio S5

S5 Prime *50 Gb*
2 x Ion 550 Chip

S5+ *30 Gb*
2 x Ion 540 Chip

S5 *15 Gb*
1 x Ion 540 Chip

200-400bp



Ion PGM

50-100 Mb
314v2 chip

600-1200 Mb
316v2 chip

1000 - 1200 Mb
318v2 chip

200-400bp



Ion Proton

15 Gb
1 x Ion PI v3 chip
80M reads

200bp

BGISEQ-500



**Combinatorial Probe-Anchor Synthesis (cPAS)
linear isothermal Rolling-Circle Replication and
DNA Nanoballs (DNB) technology**



***PE – 100bp
two chips/run x 200Gb***

Can you pick the sequencing machine for these applications?

- Large Whole-Genome Sequencing (human, plant, animal)
- Small Whole-Genome Sequencing (microbe, virus)
- Exome Sequencing
- Targeted Gene Sequencing (amplicon, gene panel)
- Whole-Transcriptome Sequencing
- Gene Expression Profiling with mRNA-Seq
- Targeted Gene Expression Profiling
- Long-Range Amplicon Sequencing*
- miRNA & Small RNA Analysis
- DNA-Protein Interaction Analysis
- Methylation Sequencing
- 16S Metagenomic Sequencing
- Shotgun Metagenomics

<https://www.illumina.com/systems/sequencing-platforms.html>



M13 - Great Hercules
Globular

Limitations of NGS

- Platform specific limitation e.g. error rate, homopolymers, limited read-length
- the presence of repetitive sequences in the genome
- the absence of high-quality reference genome

Further Reading

1. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* [Internet]. 2016 Jun [cited 2017 Oct 10];17(6):333–51. Available from: <https://www.nature.com/nrg/journal/v17/n6/abs/nrg.2016.49.html?foxtrotcallback=true>
2. Mardis ER. DNA sequencing technologies: 2006-2016. *Nat Protocols* [Internet]. 2017 Feb [cited 2017 Oct 3];12(2):213–8. Available from: <http://www.nature.com/nprot/journal/v12/n2/abs/nprot.2016.182.html>
3. Illumina Sequencing Introduction. Available at https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

Questions?

contact info:

bhoom.suk@mahidol.ac.th